Inclusive growth and the impact of intermediate goods' productivity on economic development

Ugarte Romero, Cristian Léonardo

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES ÉCONOMIQUES ET SOCIALES

"Inclusive Growth and the Impact of Intermediate Goods' Productivity on Economic Development"

Thèse présentée à la Faculté des Sciences Économiques et Sociales de l'Université de Genève

par Cristian Leonardo UGARTE ROMERO

pour l'obtention du grade de
Docteur ès Sciences économiques et Sociales
mention économie Politique

Membres du jury de thèse :

Prof. Olivier CADOT, Université de Lausanne
Prof. Jaime DE MELO, président du jury, Université de Genève
Prof. Jaya KRISHNAKUMAR, Université de Genève
Prof. Marcelo OLARREAGA, directeur de thèse, Université de Genève

Genève, le 17 décembre de 2013
Thèse No 829

La Faculté des sciences économiques et sociales, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par là, n'émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 17 décembre de 2013

Le doyen
Bernard MORARD

Impression d'après le manuscrit de l'auteur.

A mi familia :

Isaac, Marina, María Julia,

Carlos, Edgar y Simone...

Contents

Summary

The search for prosperity that would enable particularly the poor individuals in the poorest countries to improve their living conditions has long been at the core of development economics. The contribution of this thesis to our understanding of the development process is twofold. First, it focuses on the differences in the transmission of changes in aggregate economic growth to rich and poor individuals, and its consequences for income distribution and poverty. It is well known that the benefits of spells of economic growth are not homogeneously distributed across individuals, and that they might reduce or increase the gap between poor and rich individuals, and directly contribute to the decline or rise of poverty. It is therefore important to identify policies that may share more than proportionally the benefits of growth with poor individuals. These are known as 'pro-poor' economic policies.

Second, this thesis aims to better understand the industrialization strategies that may help poor countries increase the sophistication of their production and enhance the competitiveness of their exports, thereby helping them move up along the development ladder. The success of the industrialization attempts undertaken around the world so far varies significantly. Some of the failures in this area are due to the fact that industrial competitiveness is determined by several production activities along the value chain, and economic policies which are not necessarily designed to support them.

The first chapter of this thesis seeks to understand the impact of growth on poverty reduction by incorporating the standard assumption of right-skewness of the income distribution into the analysis. Dollar and Kraay (2004) propose to explore the degree of welfare transmission by means of a regression of the aggregate growth on changes of the mean income of individuals in the lower quintile of the distribution. However, this regression suffers from misspecification because it neglects the effect of distributional changes on the income of the poor. It therefore does not properly capture the transmission of economic growth to individuals of a country. Using a theoretical framework consistent with the right-skewness of incomes, we show that changes of the median income of the distribution are proportionally distributed across individuals, irrespectively of changes in inequality. On the other hand, changes of the mean income are unevenly distributed, due to the accompanying change in inequality. Based on these results, we propose an analytical framework to assess the impact of economic policies on poor individuals' income and to evaluate the likelihood of these policies to be considered 'pro-poor' or not.

The empirical evidence obtained through simulations of changes for right-skewed distributions –which are commonly used to describe the distribution of income– corroborates our theoretical results. There is a bias in the estimation of the elasticity of mean incomes of the poor on mean incomes because this does not control for dispersion or inequality in the regression. Our simulation results also show that there is no bias in the estimation of this elasticity for median incomes and, therefore, the estimated elasticity is not statistically different from its expected value, regardless of changes in income inequality. Given the bias in mean incomes due to misspecification, the above estimation procedure does not generate estimates that could be used to assess the 'pro-poorness' of income changes and of the policies that have led to these changes. However, the unbiasedness of the estimate based on median incomes allows us to discard the use of this regression framework to capture the degree of transmission, and to use instead a simplified analysis where the impact of economic policies on mean and median incomes of the population is informative enough of the pro- or anti-poor implications of policies. In fact, we show

that policies having a positive and larger impact on the median income of the population compared to their impact on the mean income are more beneficial to the poor. Given that this simplified analysis takes into account changes in central tendencies as well as changes in redistribution, it is indeed half way between the macro analysis of the impact of economic policies on aggregate measures of poverty and the micro analysis which considers the impact of policies on welfare at the household level.

The subsequent two chapters are related to the work by Jones (2011) which proposes an analytical framework for the analysis of economic development of countries, integrating complementarities and linkages across economic sectors. His model provides a formulation of the gross domestic production where productivity composites of final and intermediate goods' consumption have a direct impact on the level of development. In particular, low productivity in particular sectors, which are called *Weak Links*, can reduce aggregate welfare.

In the second chapter, we test Jones' proposition. To this end, we build an empirical measure of the probability of observing *Weak Links* by identifying sectors with low productivity which tend to be non-tradeable and whose products are heavily used as intermediate inputs by other sectors. The estimation of the impact of *Weak Links* on growth shows that their presence significantly lowers the annual growth rate of countries. Our empirical evidence also shows that the lower is the productivity of *Weak Link* sectors, the lower is economic growth, and the more statistically significant becomes the impact of those *Weak Links* on economic growth.

The last chapter assesses the impact of *Weak Links* on the relationship between diversification and development of developing countries. An important literature has shown that the relationship between economic diversification and income per capita is non-monotonic (Imbs and Wacziarg, 2003, and Koren and Tenreyro, 2007). At early stages of development, as income increases and new economic opportunities emerge, countries diversify. However, at later stages of development, as income rises beyond a certain threshold, the production bundle becomes more concentrated. The aim of this chapter is to explore the role played by *Weak Links* à la Jones (2011) in explaining the non-monotonic relationship between income per capita and economic diversification. The results of this empirical exercise show that economies where *Weak Links* are more likely to be observed tend to have a more concentrated production bundle. Moreover, the inverted u-shape relationship between income per capita and economic diversification tends to be stronger in countries where *Weak Links* are more likely to be observed.

To conclude, the findings in this thesis underline the need for comprehensive and inclusive industrial policies in developing countries. Policymakers need to have in-depth knowledge of the structure and linkages of the economy to formulate and implement informed development strategies. The key message of the second part of the Thesis is that industrial policy cannot only be targeted at a specific sector. There is a need to study the whole value chain of production and to address productivity bottlenecks, or *Weak Links*, across the sectors in order to promote sustainable and effective development. This type of interventions are more likely to have a higher and longer-lasting impact on aggregate growth and, at the same time, facilitate the transition from subsistence and extractive activities with low profit margins to high value-added industries. Moreover, trade policy can be used by policymakers to foster industrial production by increasing the tradeability of low productivity segments of the value chain following a careful analysis of the benefits and consequences of trade liberalization in internal and external markets.

Résumé

La quête de la prospérité qui permettra aux individus les pauvres d'améliorer leurs conditions de vie est un sujet au cœur de l'économie du développement. La contribution de cette thèse à cette littérature est constituée de deux parties. Premièrement, elle se concentre sur la transmission de la croissance agrégée entre les individus pauvres et riches ainsi que sur leur conséquence en termes de distribution de revenu et pauvreté. Il est reconnu que les bénéfices de la croissance économique ne sont pas homogènement distribués aux individus et par conséquent, ils peuvent réduire ou augmenter l'écart entre les individus riches et pauvres. Ainsi, ces bénéfices ont une influence directe sur l'augmentation ou la diminution du niveau de pauvreté. Il est donc très important d'identifier les politiques économiques qui permettent aux pauvres de bénéficier plus que proportionnellement de la croissance agrégée. Ce sont les politiques dites favorables au pauvre ou 'pro-poor'.

Deuxièment, cette thèse cherche à mieux comprendre les stratégies d'industrialisation qui peuvent engendrer une amélioration de la sophistication de la production des pays pauvres et ainsi accroître la compétitivité de leurs exportations. Les succès de différentes tentatives d'industrialisation varient significativement à travers le monde et quelques-uns des échecs dans ce domaine sont dus au fait que la compétitivité est déterminée par nombreuses activités productives en chaîne et que les politiques économiques ne s'adaptent pas nécessairement aux besoins de chacune de ces activités.

Le premier chapitre de cette thèse analyse l'impact de la croissance sur la pauvreté en intégrant l'hypothèse d'une asymétrie à droite de la distribution des revenus dans l'analyse. Dollar et Kraay (2004) suggèrent d'étudier le degré de transmission du bien-être en utilisant une régression du revenu moyen des individus dits pauvres sur le taux de croissance du revenu par tête. Or, cette régression n'est pas correctement spécifiée puisqu'elle ignore la redistribution de revenus qui a un impact direct dans le revenu du pauvre et par conséquent, elle n'est pas apte pour quantifier le degré de transmission des changements dans le bien-être agrégé. En outre, nous montrons que les changements du revenu médian sont proportionnellement distribués entre les individus sans être influencés par la redistribution des revenus et sur la base de ce résultat, nous proposons un nouveau cadre analytique pour l'évaluation des politiques économiques comme étant favorables au pauvre.

L'évidence empirique obtenue à travers de simulations sur des distributions asymétriques à droite confirme nos résultats théoriques et nous détectons un biais dans l'estimation de l'élasticité du revenu moyen du pauvre par rapport au revenu moyen de la population qui est dû à l'omission des termes liés à la dispersion ou inégalité des revenus. En travaillant avec des revenus médians, nos résultats ne montrent pas de biais significatifs pour cette élasticité. Cette propriété nous permet de proposer un cadre simplifié où l'analyse de l'impact des politiques économiques sur les revenus moyen et médian est suffisante pour les décrire comme favorables au pauvre ou pas. Ainsi, une politique économique avec un impact positif et plus large sur le revenu médian que sur le revenu moyen sera favorable au pauvre. Cette analyse qui prend en compte les changements en tendance centrale ainsi qu'en redistribution de revenus est en fait à mi-chemin entre l'impact sur des mesures agrégées de pauvreté et l'analyse microéconomique au niveau du ménage.

Les deux autres chapitres de cette thèse sont étroitement liés au travail de Jones (2011) qui propose un cadre analytique pour l'analyse du développement économique des pays en intégrant la complémen-

tarité et connectivité entre secteurs. Son modèle permet de dériver une expression pour la production totale où deux agrégats de productivité, l'un sur des biens finaux et l'autre des biens intermédiaires, ont un impact direct sur le niveau de développement du pays. La basse productivité de certains secteurs, appelés 'Weak Links', réduisent considérablement le niveau de bien-être.

Le deuxième chapitre évalue directement ce postulat et construit à cette fin une mesure de la probabilité d'observer des 'Weak Links' en les identifiant comme des secteurs à basse productivité, peu échangeables et intensément utilisés en tant que biens intermédiaires. Nos résultats montrent que leur présence diminue considérablement le taux de croissance annuel au niveau agrégé et que leur impact devient de plus en plus important lorsque ces secteurs deviennent de moins en moins efficients.

Le dernier chapitre analyse l'impact des 'Weak Links' sur la relation entre diversification économique et niveau de développement des pays. Une importante littérature a montré que la relation entre diversification et revenu par tête n'est pas monotone (Imbs et Wacziarg, 2003, et Koren et Tenreyro, 2007). Pour des revenus faibles, lorsque le revenu s'accroît et des nouvelles opportunités économiques émergent, les pays tendent à se diversifier. Mais au delà d'un certain seuil du revenu per capita, la concentration économique tend à augmenter lorsque le revenu s'accroît. L'objectif de ce chapitre est d'explorer l'apport des 'Weak Links' à la Jones (2011) pour expliquer cette relation non monotone entre revenu et diversification. Les résultats empiriques montrent que les économies avec plus de chances d'observer des 'Weak Links' ont une production plus concentrée et que la relation en U inversée est plus forte pour ces pays.

Pour conclure, les résultats de cette thèse soulignent le besoin de formuler de politiques industrielles cohérentes et inclusives dans les pays en voie de développement. Pour cela, il faut une connaissance approfondie de la structure ainsi que des interrelations entre secteurs économiques afin d'implémenter des stratégies réussies de développement. La politique industrielle des pays ne peut pas être axée uniquement sur un secteur spécifique mais il est nécessaire d'étudier toute la chaîne productive et de résoudre les goulots d'étranglement de la production afin de promouvoir un développement effectif et durable. Ce type d'interventions aura un impact significatif et durable sur la croissance agrégée et permettront une transition vers des activités à valeur ajoutée plus importante. La politique commerciale extérieure peut aussi servir à ces fins en permettant d'accroître la commercialisation/négociabilité des segments à basse productivité suite à une analyse complète de bénéfices et conséquences de la libéralisation commerciale de ces secteurs.

Remerciements

Cette thèse n'aurait pas été possible sans la contribution d'un très large nombre de personnes et de situations au cours des treize dernières années à Genève. Ces quelques lignes ne sont qu'une petite reconnaissance de ma part envers eux et il s'agit certainement d'une liste incomplète de personnes à qui je dois remercier pour leur soutien, écoute et encouragement tout au long de mes études.

D'abord, je dois remercier à l'Université de Genève puisqu'elle m'a accueilli dès le début de mes études universitaires en 2001 et il s'agit aujourd'hui de la fin d'un processus étroitement lié au Département de Sciences Économiques auquel je suis entièrement reconnaissant dans son ensemble pour la formation que j'ai reçu. Je tiens à remercier particulièrement à plusieurs professeurs qu'à différents moments de mon cursus ont su m'encourager pour arriver à la fin de cette thèse. Je dois ainsi remercier à Gérard Antille, Fabrizio Carlevaro, Manfred Gilli et Elvezio Ronchetti pour leur soutien au cours de mes études, à Jaya Krishnakumar sans qui cette thèse n'aurait jamais débuté, à Tobias Müller pour des longues conversations académiques et non-académiques, à Céline Carrère pour ses encouragements et son soutien desintéréssés, à Jaime de Melo pour avoir pris soin de moi et de ma thèse pendant ces dernières années et bien entendu à Marcelo Olarreaga sans qui j'aurai quitté ce projet à plusieurs reprises mais qui a toujours améné de l'enthousiasme et de la bonne humeur en plus de sa connaissance pour me donner de l'élan dans mon travail. Je remercie aussi au Professeur Olivier Cadot qui fait partie du jury en tant que membre externe mais qui a donné des commentaires sur ces travaux dès les prémières esquisses.

Treize ans sont une longue période de temps et donc je n'aurai pas pu survivre à cet épreuve sans le soutien de plusieurs amis et je suis très chanceux de les avoir rencontré dans mon chemin. J'espère que personne ne tiendra compte de ma mémoire non exhaustive mais je dois dire merci à Alan, Alfredo, Alejandro, Andrés, David, Krista, John, Laura, Lucas, Pablo, Wayra et bien d'autres. Un grand merci en particulier à Patricia et Raquel sans qui je serai rentré dans mon pays avant la fin de ma Licence à Genève, à Jorge pour ses interminables conversations dans les couloirs d'Uni Mail, à Javier pour ces longues et très amusantes années ensemble à Genève et à Daniel pour son amitié inconditionnelle et pure.

Je leur dois toute ma gratitude à ceux qui ont eu à souffrir directement de mes divagations pseudo-intelectuelles dans ce Doctorat et je remercie à tous mes collègues au Département : Alain, Liliana et Pramila pour cette expérience académique (hebdomadaire) qui restera en moi pour toujours, Juan(cho) sans qui le DEA aurait été insurmontable, Laura pour cette amitié qui va au delà de l'académique et mille mercis à Anjeza, Brahim, David, Gerda, Giovanni, Julie, Thierno et bien d'autres.

Finalement, j'ai bénéficé d'un soutien inconditionnel de la part de ma famille qui a accepté de sacrifier de vacances et des moments ensemble pour que je puisse me dévouer plus à mon travail. Merci à mes parents qui ont fait des efforts hors norme pour mon éducation, à mes frères et ma soeur pour leur écoute continue de mes malheurs et bonheurs et merci à toi, Simone, qui est la grande tributaire de cette thèse de Doctorat sans autre intérêt que le mien.

Chapter I

The Effect of Growth on Poverty

CHAPTER I
The Effect of Growth on Poverty

Jaya Krishnakumar and Cristian Ugarte[*]

Department of Economics

University of Geneva, Switzerland

Abstract

The aim of this paper is to understand the impact of growth on poverty reduction in a right-skewed income distribution setting. Existing literature focuses on the effect of aggregate growth on changes of the mean income of the poor by means of a regression of the latter on the former. This, however, neglects the effect of other distributional changes (e.g. inequality change) on the income of the poor thus making the above regression misspecified. This paper proposes a suitable theoretical framework consistent with the right-skewness for examining the growth impact on poverty taking inequality (dispersion) into account. We show that an economic policy that positively affects the *median* income of the distribution has a proportional change on the *median* income of the poor, whatever the change in inequality. On the other hand, when a policy has a positive impact on the *mean* income, its effect on the *mean* income of the poor can be uncertain depending on the accompanying change in inequality. This is illustrated by simulations as well as an empirical application.

Keywords: Pro-poor growth, Poverty, Inequality, Income distribution.

[*]We are grateful to Olivier Cadot, Jorge Davalos, Jaime de Melo, Marcelo Olarreaga and Laura Zoratto for very helpful discussions as well as participants at the CUSO doctoral workshop in Economics and Econometrics and the Young Swiss Economists Meeting 2010. All remaining errors are the authors' responsibility. E-mail: jaya.krishnakumar@unige.ch, cristian.ugarte@unige.ch .

Contents

List of Tables

List of Figures

1 Introduction

Literature on whether economic growth is pro-poor or not is fast expanding especially in recent years. Public debates associated with aggregate economic reforms often include discussions over the likelihood of policy changes to promote pro-poor growth[1]. An important part of this literature first identifies a poverty line and then verifies whether aggregate growth or economic reforms enable poor households to raise their income above this threshold or how these policies reduce the likelihood of observing poor individuals (Adams, 2004, Bourguignon, 2004a, Chen and Ravallion, 2001, Ravallion and Datt, 2001, Sala-i-Martin, 2006)[2]. An alternative approach explored by Dollar and Kraay (2002, 2004) and others is to see how changes in aggregate income affect the mean income of the poor[3]. If the elasticity of the mean income of the poor with respect to aggregate income is larger than 1, we can argue that aggregate growth is pro-poor. This is the approach followed in this paper.

We argue that the regression using mean incomes is not able to disentangle the impact of different changes that occur in growth and their impact on the income of the poor. There are two different aspects to aggregate growth: one associated with increases in the central tendency of the distribution (for a given dispersion), and the other one associated with changes in the dispersion (for a given mean/median income)[4]. We show that the former is relative-poverty-neutral in the sense that as median income increases, income of the poor increases proportionately provided that the dispersion does not change. The latter, however, is key to pro- or anti-poor growth.

The objective of this paper is twofold. First to theoretically derive the elasticities of the mean and median incomes of the poor with respect to those of the population assuming a right-skewed income distribution and examine whether a regression with mean incomes as usually done is able to capture the true values of the parameters of impact of aggregate growth on the income of the poor. Second, to derive a methodology

[1]The 'pro-poorness' of a policy is closely related to the definition of 'pro-poor' that is considered. In this context it is relevant to distinguish between absolute and relative pro-poorness. Absolute pro-poorness describes a positive impact on the level of income of poor individuals while relative pro-poorness indicates that poor individuals obtain a larger share of the benefits. For further details, see Klasen (2008).

[2]The 'growth elasticity' of poverty initially proposed by Kakwani (1993) is the key empirical issue in this strand of the literature and is given by the impact of growth on a poverty indicator such as the headcount or gap ratio.

[3]Some authors have examined this question using household survey data by estimating the effect of aggregate growth on different income levels. The poverty growth curve (PGC) proposed by Son (2004) and the growth incidence curve (GIC) by Ravallion and Chen (2003) are two important attempts to characterize pro-poor growth spells. Duclos (2009) is another important contribution on this topic that relies on dominance methods for evaluating the pro-poorness of changes in income distributions

[4]These aspects have been taken into consideration by authors such as Adams, 2004, Bourguignon, 2004b, Kakwani and Pernia, 2000, Kraay, 2006, Ravallion, 2001, Ravallion and Huppi, 1991, who have each followed a different strategy to evaluate the impact of growth.

to assess the impact of economic reforms on the income of the poor by comparing the two channels identified above (mean and median incomes) thus enabling the quantification of the trade-offs involved in poverty reduction policies as suggested in Ravallion (2004).

As an example, trade reforms will affect the poor through their impact on median income, but also through the dispersion of the income distribution. Indeed the well known Stolper-Samuelson theorem tells us that trade reforms change relative factor prices which will lead to changes in factor and income inequality. These two channels were identified in Dollar and Kraay (2004) who explored the impact of trade on poverty in three steps. First, Dollar and Kraay (2004) estimated the impact of trade on aggregate growth and found a positive relationship. Then, they explored the relationship between aggregate mean income and the mean income of the poor, and found a one to one relationship (as in Dollar and Kraay 2002). Finally, because trade also affects the dispersion of income, they explored whether trade leads to more or less income inequality, and they found no statistically significant impact[5]. These three steps put together allowed them to conclude that there is a positive impact of trade reforms on the mean income of the poor. We argue that the methodology used to estimate the impact of growth on the income of the poor is not suitable to inform on the 'pro-poorness' of growth spells and its estimates tend to be downward biased except if there are no changes on inequality[6].

Based on Dollar and Kraay (2002), we derive estimates of the impact of growth on the income of the poor within a framework that is theoretically consistent with right-skewness of the income distribution and that simultaneously takes into account the impact of both aspects of growth - increase in mean/median income and change in variance/dispersion. We test the accuracy of commonly used regressions to capture estimates of this impact and finally, we propose a methodology that is midway between the analysis of the impact of policies using household data and the analysis of the impact of aggregate growth on poverty (Bourguignon, 2004a, Ravallion, 2001). Our proposal is to study simultaneously the impact on mean and median incomes in order to conclude on the 'pro-poorness' of a policy or shock and given this dual analysis, the proposed method considers distributional impact of policies jointly with the usual impact on aggregate growth. Given right-skewness of income distribution, the evolution of these two measures would be sufficient to describe the distributional impact of a policy affecting all individuals.

The rest of the paper is organized as follows. Section 2 presents two measures of average income under the assumption of right-skewness. Section 3 explores the adequacy of the different measures of growth or income level to explain the average growth of income of the poor or its level. Section 4 presents results obtained using simulations and US

[5] A more comprehensive evaluation of how trade liberalization affects income inequality can be found in Gourdon, Maystre and de Melo (2008).

[6] In case inequality remains constant, the empirical exercise is trivial as the change in income for all individuals is proportional.

income data on the accuracy of estimates of the elasticity of the income of the poor with respect to growth for different measures. Section 5 derives the absolute impact of changes in central tendency and dispersion of income on poverty reduction, under the assumption of right-skewness. Section 6 concludes.

2 Defining measures of incomes

In order to study the impact of growth on poverty one first needs to select appropriate measures of income for the whole population and that for the poor and then investigate the relationship between the two. In general, the literature has taken the mean incomes of the whole population and the mean incomes of the poor as the appropriate measure for looking at the impact of growth on poverty. However, some other measures such as the median which also characterize the central tendency of a distribution can be used. In fact, median measures turn out to be better in practice in the case of extreme observations in the empirical distribution because they are less sensitive to extreme values. We will therefore consider both the mean and the median for our investigation.

Mean income of the whole distribution will serve as our yardstick for growth and poverty will be measured by that of the poorest quintile of the distribution. Instead of defining a poverty line[7], we prefer to follow Dollar and Kraay (2002, 2004) and work on the relative evolution of incomes instead of measuring the impact of aggregate growth on a set of poverty indicators. The use of poverty indicators is informative when our interest is focused on the incidence of poverty but it lacks to inform researchers on the evolution of relative differences in income levels as we will not be able to disentangle how growth benefits are distributed across the entire distribution. Hereafter, we define the poor to be individuals whose incomes are contained in the bottom quintile of the distribution. Comparisons of the evolution of incomes would shed some light on the transmission of growth within the distribution.

We derive our theoretical results for two commonly fitted income distributions : lognormal and Pareto[8]. Denoting income as y, the theoretical mean income is the first

[7]In fact, it has been shown that the 'growth elasticity' of poverty obtained using different indicators of poverty might be highly sensitive to the location of the poverty line in the income distribution (Heltberg, 2002, and Ravallion and Datt, 2001) and also to the initial level of inequality (Ravallion, 2001).

[8]The suitability of these distributions to describe incomes depends on which part of the distribution our interest is focused. The lognormal distribution fits lower incomes better than the Pareto distribution but it is less satisfactory towards the upper end compared to the Pareto distribution (Lombardo, 2009). Lopez and Serven (2006) and Klasen and Misselhorn (2008) provide evidence on the empirical fit of lognormal distributions to income data. Lopez et al. cannot reject the null hypothesis that per capita income follows a lognormal distribution. Klasen et al. show that estimated characteristics derived from a lognormal distribution fits properly the observed values of absolute poverty calculated with income data.

order moment of the distribution ($E(y)$) and its empirical equivalent is the income per capita. The median income is the observation minimizing the absolute distance to all other observations of the income distribution and it corresponds to the argument c in the following mathematical expression :

$$M(y) = \arg\min_{c} E(\mid Y - c \mid). \tag{1}$$

In what follows, we will denote the theoretical mean as $\overline{y}$ and $M(y)$ as y^M. Even if lognormal and Pareto distributions are commonly used to describe incomes, our choice is not based only on their empirical suitability. Both distributions allow us to obtain mathematical expressions for the general (aggregate) measures of income and for the measures restricted to the bottom quintile. The lognormal distribution can be interpreted as a monotonic transformation of the Normal distribution, an important property that will be used while working in the bottom quintile. It is well defined for positive values and is right-skewed as an income distribution generally is. The Pareto distribution specified with a lower bound greater than zero satisfies similar properties.

Lognormal distribution

Under log-normality, the mean income of the distribution is given by the parameters of the underlying Normal distribution $N(\mu;\sigma^2)$. The median income of the distribution can be obtained from the median value of the associated normal distribution. We will note these two measures respectively $\overline{y}^{LN}$ and $y^{M,LN}$. Their expressions are given by :

$$\overline{y}^{LN} = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad , \quad y^{M,LN} = \exp(\mu). \tag{2}$$

Let us now work out the expression of the two analogous measures for the individuals in the bottom quintile, in other words the mean income ($\overline{y}_p$) and the median income of the same quintile (y_p^M)[9].

The mean income over a defined interval is the conditional expectation of the distribution on this interval. The expression of this expectation is[10]:

$$\overline{y}_p = \int_{Q1} y f_{Q1}(y \mid y \in Q1) dy = 5 \times \int_{Q1} y f(y) dy, \tag{3}$$

[9] The choice of the bottom quintile should not be interpreted as a poverty line since we are interested in the evolution of an 'average' income within the bottom quintile and not in the proportion of individuals below the upper bound of the bottom quintile. In fact, this proportion is constant and will always be equal to 0.2.

[10] See Appendix A.1 for further calculations.

where $f_{Q1}(\cdot)$ is the conditional density function for the interval $Q1$, the bottom quintile of the distribution, and $f(\cdot)$ is the unconditional density function. The interval $Q1$ is defined by $[0^+, t_{0.2}]$, where $t_{0.2}$ is the 20% percentile of the distribution.

Given that a lognormal distribution is a transformation of a Normal distribution, we use the upper bound of the bottom quintile of the Normal distribution $N(0,1)$, $z_{0.2}$, and the definition of the mean of a truncated lognormal distribution to obtain the mean income of the poor[11] :

$$\overline{y}_p^{LN} = 5 \times \exp\left(\mu + \frac{\sigma^2}{2}\right) \times [1 - \Phi(-z_{0.2} + \sigma)] \tag{4}$$

By definition, the median income of the poor is the 10 % percentile of the distribution $f(\cdot)$. Under log-normality, it is given by :

$$y_p^{M,LN} = \exp(\mu + z_{0.1}\sigma) \tag{5}$$

Pareto distribution

Now, let us turn to the Pareto distribution whose density function is the following :

$$f(y) = \frac{k y_m^k}{y^{k+1}} \qquad \text{with} \quad y_m > 0 \quad \text{and} \quad k > 1 \tag{6}$$

where y_m is the lower possible value of the distribution and the restriction over the parameter of dispersion k is necessary for the existence of the mean value of the distribution (income per capita). The mean and the median of the distribution are given by :

$$\overline{y}^P = \frac{k y_m}{k-1} \qquad \text{and} \qquad y^{M,P} = y_m \times 0.5^{-\frac{1}{k}} \tag{7}$$

In this case, the mean income and the median income of the poor in the bottom quintile are defined as follows[12] :

$$\overline{y}_p^P = 5 \times \frac{k y_m}{k-1} \times \left[1 - 0.8^{1-\frac{1}{k}}\right] \tag{8}$$

$$y_p^{M,P} = t_{0.1}^P = y_m \times 0.9^{-\frac{1}{k}} \tag{9}$$

[11]See Appendix A.1 for further calculations.
[12]Idem.

3 How well are we doing by regressing $\overline{\mathrm{y}}_{\mathrm{p}}$ ($\mathrm{y}_{\mathrm{p}}^{\mathrm{M}}$) on $\overline{\mathrm{y}}$ (y^{M}) ?

If we introduce the logarithm of an overall measure of income as an explanatory factor in an equation for the logarithm of the measure of income for the poor, then the coefficient of the former can be interpreted as an 'elasticity' of the mean (median) income of the poor to growth measured as a percentage increase in the mean (median) income of the population. Hence a tempting idea would be to run a log-log regression where the explained variable is the logarithm of one measure of income for the poor (m_p) with the explanatory variable as the logarithm of the same measure of income for the population (m_D). Such a regression provides an estimate of the relative variation of the income for the poor as the income of the population changes. This would give the following formulation :

$$\ln m_p = \alpha_0 + \alpha_1 \ln m_D + \varepsilon. \tag{10}$$

In a growing economy, if $\hat{\alpha}_1 > 1$ growth would be 'pro-poor'. The case $\hat{\alpha}_1 = 1$ could be interpreted as a 'neutral' evolution of incomes with the income share of the poor remaining unchanged. The case $\hat{\alpha}_1 < 1$ implies that an increase in the overall income is associated with a less than proportional increase of low incomes.

In what follows we show that the above relationship can be misspecified from a theoretical point of view depending on which measure of income is chosen. We successively consider the mean and the median and examine the theoretical relationship.

First, let us take the mean income. If incomes were distributed according to a lognormal distribution, we have :

$$\ln m_p = \ln \overline{y}_p^{LN} = \left(\mu + \frac{\sigma^2}{2}\right) + \ln\left[1 - \Phi(-z_{0.2} + \sigma)\right] + \ln(5) \tag{11}$$

and

$$\ln m_D = \ln \overline{y}^{LN} = \mu + \frac{\sigma^2}{2} \tag{12}$$

Thus, under log-normality, regressing $\ln \overline{y}_p$ on $\ln \overline{y}$ as in Equation 10 should lead us to $\hat{\alpha}_1 \approx 1$ provided the omitted term $\ln\left[1 - \Phi(-z_{0.2} + \sigma)\right]$ (let us call it the dispersion term $disp(\sigma)$ as it involves σ) is uncorrelated with $\overline{y}^{LN}$. In this case the OLS estimator would be unbiased. In Figure 1a we plot this term against σ to see the relationship between the term and σ for the interval chosen. The boundaries of the interval were calculated from the observed Gini values for countries across the world (between 0.25 and 0.8) and using the relationship between Gini index and σ (see Appendix A.2). Thus σ ranges from 0.45 to 2. It is important to understand the behavior of $disp(\sigma)$ with respect to σ which would be crucial in further calculations. Higher values of σ imply higher inequality in the income distribution and consequently, this reduces the income share of the bottom quintile. This is reflected in Figure 1a where $disp(\sigma)$, the log of the income share of the bottom quintile, is a decreasing function of σ (and income inequality).

[FIGURE 1 HERE]
[DISPERSION TERMS OF LOG-MEAN INCOMES FOR SELECTED DISTRIBUTIONS]

Similarly, for the Pareto distribution, we have

$$\ln m_p = \ln \overline{y}_p^P = \ln \left(\frac{k y_m}{k-1} \right) + \ln \left[1 - 0.8^{1-\frac{1}{k}} \right] + \ln(5) \tag{13}$$

$$\ln m_D = \ln \overline{y}^P = \ln \left(\frac{k y_m}{k-1} \right) \tag{14}$$

and the unbiasedness of the OLS estimator of α_1 will depend on the covariance between the logarithm of $\overline{y}^P$ and $\ln[1 - 0.8^{1-\frac{1}{k}}]$. Here the dispersion term depends on k which ranges from 1.125 and 2.5 according to the observed different values of Gini index (see Appendix A.2) and we call it $disp(k)$. Figure 1b traces the dispersion term for different values in this interval. At first glance, the behavior of the dispersion term for the Pareto distribution seems to be opposite to the result in Figure 1a. Therefore, it is important to note that lower values of k are associated with fat tails on the right of the income distribution and therefore, they imply higher inequality. The log of the income share of the bottom quintile, $disp(k)$, is thus an increasing function of k but a decreasing function of income inequality.

Taking a linear approximation for both dispersion terms i.e. $disp(\sigma) \approx a + b\sigma$ and $disp(k) \approx a + bk$, we can calculate an expression of the bias concerning the estimation of α_1 in Equation 10 as shown in Equations 15-17. The bias of the estimation shown below is related to the omission of the dispersion term and including a function of the variance of the distribution (e.g. inequality measure) will certainly correct for the bias but will yield $\hat{\alpha}_1 \rightarrow 1$ making the empirical exercise redundant when the empirical distribution is close to the ones assumed here.

When using the mean values of income, the estimate of α_1 obtained by ordinary least squares for the lognormal distribution would be:

$$\hat{\alpha}_1 = \frac{Cov(\ln \overline{y}; \ln \overline{y}_p)}{V(\ln \overline{y})} = \frac{E\left[\ln \overline{y}_p \ln \overline{y}\right] - E\left[\ln \overline{y}_p\right] E\left[\ln \overline{y}\right]}{E[(\ln \overline{y})^2] - E^2\left[\ln \overline{y}\right]}. \tag{15}$$

Replacing $\ln \overline{y}_p$ by $\ln \overline{y} + a + b\sigma$, $\hat{\alpha}_1$ can be written as :

$$\hat{\alpha}_1 \approx 1 + b \quad \left[\frac{Cov\left(\mu, \sigma\right) + Cov\left(\frac{\sigma^2}{2}; \sigma\right)}{V[\ln \overline{y}]} \right] \tag{16}$$

The bias of $\hat{\alpha}_1$ will be proportional to $Cov(\mu;\sigma) + Cov\left(\frac{\sigma^2}{2};\sigma\right)$. $Cov\left(\frac{\sigma^2}{2};\sigma\right)$ is positive (see Appendix A.3). We now make an additional assumption on the characterization of the parameters of the distribution, μ and σ and we assume $Cov(\mu;\sigma) = 0$. Thus assuming $Cov(\mu;\sigma) = 0$, we have that, for any size of the sample, the estimate $\hat{\alpha}_1$ is biased, with the bias taking the sign of b i.e. negative.

Even though the assumption of no covariance between central tendency and dispersion across countries could be seen as too strong, it is compatible with the well-known Kuznets' hypothesis that suggests a concave relationship between inequality and income levels. Given the concavity of this relationship, covariance would be on average very close to zero. Indeed no covariance implies no linear relationship on average while in reality, this relationship might be positive or negative within some countries. Besides that, it is also true that the assumption of no covariance hides the intuition in Dollar and Kraay (2002) which actually seeks to capture this correlation. Nonetheless, it has one practical advantage which is to set a natural benchmark for the results of any estimation, $\hat{\alpha}_1 = 1$. Otherwise, we would not have a direct comparison for the theoretical results and the theoretical bias in Equation 16 would remain the same but just with respect to $1 + b \times Cov(\mu,\sigma)/V[\ln \overline{y}]$. Note that in case $Cov(\mu;\sigma) > 0$, the theoretical estimate of α_1 could still be lower than 1 while the expected estimate is higher than this value and this is against the intuition presented in Dollar and Kraay (2002).

A similar expression holds for the Pareto distribution assuming $disp(k)$ to be linear on k. The estimate of α_1 is given by :

$$\hat{\alpha}_1 \approx 1 + b \left[\frac{Cov(\ln\left(\frac{k}{k-1}\right), k)) + Cov(\ln y_m, k)}{V[\ln \overline{y}]} \right] \tag{17}$$

where $Cov(\ln(\frac{k}{k-1}), k)$ is negative (see Appendix A.3). Assuming again that there is no covariance between tendency and dispersion, $Cov(\ln y_m, k) = 0$, the bias of $\hat{\alpha}_1$ will take the opposite sign of b, i.e. negative. Thus we conclude that the regression on means should produce estimates of the elasticity coefficient lower than 1. The justification for $Cov(\ln y_m, k) = 0$ are the same as those presented before for $Cov(\mu, \sigma) = 0$.

On the other hand, using the median as our indicator of income we see that regressing the median income of the poor on the median income of the population will not yield a biased estimator under log-normality, maintaining the assumption $Cov(\mu;\sigma) = 0$. For the Pareto distribution, the theoretical estimate is however less than one (see Appendix A.4) :

$$\hat{\alpha}_1^{M,P} = 1 + \ln(0.5) \times \ln\left(\frac{9}{5}\right) \times \frac{V\left(\frac{1}{k}\right)}{V(\ln y^{M,P})}. \tag{18}$$

Table 1 presents the summary of theoretical estimates that would be observed in case $Cov(\mu, \sigma) = 0$ holds, on average, for the lognormal distribution and in case $Cov(\ln y_m, k) = 0$ holds, on average, for the Pareto distribution. Transmission of changes in median incomes is close to be proportionally distributed across the population while changes in mean incomes fail to be homogeneously distributed.

[TABLE 1 HERE]
[SUMMARY OF BIAS IN $\hat{\alpha}_1$]

4 Simulations and Empirical Illustration

4.1 Simulations

In this section, we present the results of simulations carried out separately for both lognormal and Pareto distributions. For the lognormal distribution, a random sample of 200 pairs (μ, σ) uniformly distributed on $[4; 9] \times [0.45; 2]$ was created. The range of values for σ is derived from the relationship between the Gini index and the variance under lognormality[13] as described in Appendix A.2. For each pair, a distribution of income for a population of 1000 observations was generated. For each simulated population (distribution), we extracted the empirical values of $\overline{y}^{LN}$, $\overline{y}_p^{LN}$, $y^{M,LN}$ and $y_p^{M,LN}$and ran the regression specified by Equation 10 with both the mean and median as measures of income. The results are reported in the first panel of Table 2 below.

[TABLE 2 HERE]
[REGRESSION OF THE MEASURE OF INCOME OF POOR ON THE MEASURE OF INCOME OF THE POPULATION]

This Table shows that when mean incomes are employed as measures, the coefficient estimates of α_1 are not 'close' to 1, the true value. The same conclusions can be drawn when the changes on this measure are used in the regressions. Estimates in regressions using mean incomes are different from 1 at usual levels of statistical significance.

[13]There is no formal justification for the range of values for μ.

Further, looking at the R^2 values for these regressions, one can see that both levels or changes of the mean income of the population have a lower explanatory power in explaining changes or levels of the mean income of the poor[14]. When median incomes are used, the estimate is not significantly different from 1, the true value, and this assumption cannot be statistically rejected.

Results in the second panel of Table 2 correspond to simulations done with the Pareto distribution with values randomly generated for k and y_m over the space $[1.125; 2.5] \times [1'000; 5'000]$. Here again, the range of values for k is determined by the relationship between the Gini index and the dispersion parameter as described in Appendix A.2[15]. The estimates for the Pareto distribution show similar characteristics to those obtained for the lognormal distribution. We therefore conclude that increases in the median income of the population seem to induce proportional increases of the level of incomes for the poor. Increases in the mean income are not necessarily proportional for all individuals given that estimates of the elasticity are lower than 1.

Working under a particular income distribution as done above can be seen as restrictive since any observed income distribution does not need to necessarily be close to one single distribution. In order to simulate distributions that could better resemble empirically observed distributions, we decided to generate samples from a mix of the above two distributions and explore the relationship (Equation 10) using samples from the 'mixed' distributions. Our results confirm our findings with samples derived from single parametric distributions. They are presented and discussed in Appendix A.5. They clearly show that the elasticity in median incomes is larger than the elasticity in mean incomes and the former elasticity is not statistically different from 1.

4.2 Bootstrap

Next, we perform a small bootstrap analysis (recreating the sample of 200 populations 40 times) in order to study the finite sample properties of our estimates. Figure 2a presents the resulting boxplot for the estimate of α_1 in the each of the four versions of the regression for the lognormal distribution and Figure 2b for the Pareto distribution. The empirical distributions confirm our theoretical findings. Estimates using median measures of income are higher than estimates with mean incomes and nearer to 1. Regressions on mean incomes are less likely to capture the true (theoretical) value of the elasticity under the assumptions (i) that the above distributions can describe incomes and (ii) that $Cov(\mu, \sigma) = Cov(\ln y_m, k) = 0$ hold on average even though not

[14]Sensitivity to extreme values would be a potential explanation for a lower explanatory in regressions on mean values. Nonetheless, all measures considered here will be affected by extreme values except the overall median income of the lognormal distribution.

[15]Values for y_m represent the lowest income observed in each economy and we consider that the proposed range would be in line with empirical observations of these minimum values.

necessarily in every sample due to its random generation.

[FIGURE 2 HERE]
[ELASTICITY ESTIMATES FOR DIFFERENT DISTRIBUTIONS]

Figure 2c and 2d show the results obtained with two other distributions frequently used to describe incomes, the Weibull and the Singh-Maddala distributions, for which we implement the same bootstrap analysis. The ranges of values for parameters follow the same logic as in the two previous cases and they aim to replicate observed values for Gini indices. Given the expressions of mean/median incomes for the Weibull and Singh-Maddala distributions, it is not possible to clearly derive the expression of estimates as shown in Equations 16 and 17. Nonetheless, the empirical behavior of estimates for the several versions of Equation 10 allows us to confirm the results obtained with the lognormal and the Pareto distributions and we conclude that increases in median incomes are more homogenously distributed across the population than increases in mean incomes.

The assumption of no covariance between central tendency and dispersion might be seen as a restrictive and therefore, not interesting for the simulation exercise. Appendix A.6 aims at addressing doubts concerning the efficiency of the regression in mean incomes to capture 'pro-poor' spells of growth when this covariance is positive. For these simulations, increases in median incomes are negatively correlated with increases in income inequality and a doubling of median incomes implies a 5 percent reduction in income inequality, measured by the Gini index. Despite this clearly 'pro-poor' scenario, the regression in mean incomes fails to capture this characteristic for the lognormal and the Pareto distributions. On the other hand, the regression in median incomes captures is able to capture this characteristic in both cases.

4.3 Illustration with US income data and in a panel of countries (1978-2012)

Using annual data for income quintiles from the US Census Bureau for the period 1966-2009, we empirically verify the results we obtain in simulations. Mean incomes per quintile are provided and thus we can calculate the mean income of the population by addition. The results in the first column of Table 3 correspond to Equation 10 run on mean incomes. Median incomes by quintile are not available and hence we take the mean incomes of the first and third quintile as proxies for the median incomes of the poor and of the population respectively. Results in the second column are from the regression using these proxies. These results confirm our findings from simulations that changes in median income tend to be more evenly distributed within the population. Since some methodological changes were introduced in 1993 affecting the calculations of the quintile mean incomes[16], we also performed the same regressions separately for the periods before and after 1993 and the results do not change in any significant manner.

[TABLE 3 HERE]
[REGRESSION WITH US INCOME DATA]

Table 4 uses all available data on per capita income and income shares for the first and third quintiles of the income distribution from the World Development Indicators. The resulting sample includes 153 countries over the period 1978-2012 and we use this dataset to estimate the regressions on mean/median incomes following the specification in Equation 10. Here again, we approximate the median income of the population and the median income of the poor by the mean income of the third and first quintile, respectively. Three versions of each regression are run: (i) a first one with all available data, (ii) a second one with data since 2005 and (iii) a last one with the most recent observation since 2005 for all countries.

Two important results need to be highlighted here. First, the estimate of α_1 based on median incomes is consistently higher than the one based on mean incomes which implies that changes in median incomes are more homogeneously distributed than changes in mean incomes and this comparison corroborates results obtained through simulations. Second, the estimate of α_1 based on median incomes is not statistically different from one in any of the three versions of this specification. On the other hand, the estimate based on mean incomes is always lower than one at 10% significance level. These results allow us to confirm that the assumption of no covariance between central tendency and dispersion is not arbitrary, otherwise α_1 would not be close to one for regressions in median incomes.

[TABLE 4 HERE]
[REGRESSIONS IN A PANEL OF COUNTRIES WITH INCOME SHARES 1978-2012]

5 Effect of growth on poverty taking account of inequality change

When growth (measured on $\overline{y}$) occurs with a combination of changes - changes in central tendency and changes in inequality, it is interesting to ask the question about the *net* impact of this growth on the income of the poor. As the expressions for $\overline{y}_p$ and y_p^M are

[16]In 1993, the US Census Bureau increased the upper end of the income distribution that might be reported by respondents. As this change can have considerable impact on the aggregate shares used to calculate mean incomes per quintile, they suggest that measures pre-1993 and post-1992 are not directly comparable.

known, we can answer this question by deriving the total differential of these measures of income with respect to the parameters of the distribution. The total differential of the logarithms of $\overline{y}_p$ and y_p^M for the lognormal and the the Pareto distributions are given by Equations 19-22:

$$\mathrm{d}\ln \overline{y}_p^{LN} = \mathrm{d}\mu + \left[\sigma - \frac{\phi(-z_{0.2}+\sigma)}{1-\Phi(-z_{0.2}+\sigma)}\right]\mathrm{d}\sigma = \mathrm{d}\mu + \left[\sigma - h(-z_{0.2}+\sigma)\right]\mathrm{d}\sigma, \quad (19)$$

$$\mathrm{d}\ln \overline{y}_p^{P} = \frac{1}{y_m}\mathrm{d}y_m + \left[-\frac{1}{k(k-1)} - \frac{0.8^{1-\frac{1}{k}}}{1-0.8^{1-\frac{1}{k}}} \times \frac{\ln(0.8)}{k^2}\right]\mathrm{d}k, \quad (20)$$

$$\mathrm{d}\ln y_p^{M,LN} = \mathrm{d}\mu + z_{0.1}\mathrm{d}\sigma, \quad (21)$$

$$\mathrm{d}\ln y_p^{M,P} = \frac{1}{y_m}\mathrm{d}y_m + \frac{1}{k^2}\ln(0.9)\mathrm{d}k, \quad (22)$$

where $h(\cdot)$ is the hazard function of a normal distribution $N(0,1)$.

The effect of a change in the central tendency (μ or y_m) has the expected sign and affects all the individuals by the same extent. This is valid for both measures of income. In the lognormal case, the effect of σ is clearly negative for the log-median income of the poor and it depends on the hazard function $h(\cdot)$ for the log-mean income. In Figure 3 we plot the partial derivatives of log-mean income with respect to changes in the dispersion of the distribution. Figure 3a lets us conclude that the effect is also negative for the log-mean income of the poor for values of $\sigma \in [0.45; 2]$. For the Pareto distribution, a change in k has a negative impact on the log-median income of the poor given that $\ln(0.9) < 0$. Figure 3b shows the partial derivative of the log-mean income of the poor against k values. Here, an increase on k which corresponds to a reduction in inequality results in a decrease in the log-mean income of the poor.

[FIGURE 3 HERE]
[PARTIAL DERIVATIVES OF LOG-MEAN INCOME TO CHANGES IN DISPERSION]

Thus we cannot formulate a general relationship between inequality and income of the poor. In one case (Pareto), as inequality reduces (k gets bigger), the mean and the median income of the poor decreases. Such behaviour would imply a trade-off between poverty and inequality, measured as income variance. Nonetheless, Ravallion (2005) reports that there is no *empirical* support for the trade-off between poverty and inequality as a general pattern and he argues that this is possibly due to the practical absence of correlation between economic growth and changes in inequality. In the other case (lognormal), a reduction in inequality (smaller σ) implies higher levels of mean and median income of the poor and this conclusion coincides with findings in Lambert (2010) who argues that growth is pro-poor if and only if it is inequality-reducing in a lognormal world. The reason for this contradiction, summarized in Table 5 is that the behavior of these two distributions differ as inequality changes. The lognormal distribution tends to concentrate itself around the median value as σ decreases. In the case of the Pareto distribution, the observations tend to concentrate around the lower bound of the distribution (y_m) as dispersion decreases. Since the partial derivatives of $\overline{y}_p$ and y_p^M cannot be signed in the same way with respect to inequality for the two distributions, we are unable to establish a conclusion on poverty of an inequality reduction. A policy which positively affects the central tendency of the distribution (with dispersion unchanged) will definitely reduce poverty; however the net effect of a policy that affects the central tendency as well as dispersion (inequality) on poverty is uncertain being dependent on the particular distribution that characterizes the empirical observations and it is beyond the scope o this paper to establish which distribution fits the data better[17].

[TABLE 5 HERE]
[CHANGES IN POOR'S INCOME FOLLOWING REDUCTIONS IN INEQUALITY]

But our analysis does not stop here. Our theoretical framework enables us to go deeper into this issue of policy impact and deduce the pro-poor nature (or not) of a policy by making some interesting comparisons. Assuming that changes of all the measures are observed between two periods of time, we can establish the impact of a positive or negative evolution of the aggregate measures (mean and median income of the population) on poverty. Let us take the case of growth i.e. a positive change in μ under log-normality. If change in log (percentage change) of the median income of the distribution is higher than that of the mean income of the distribution, this necessarily implies that inequality has reduced i.e. $\mathrm{d}\sigma < 0$. This in turn implies that the relative changes on the mean and median incomes of the poor are respectively higher than the changes in the mean and median income of the population and we can conclude that growth has benefited more the poor (see Appendix A.7 for calculations). Relative and absolute poverty measures should have decreased. Similarly for the Pareto distribution, if the percentage change on median income is higher than the percentage change on the mean income, it means dk is positive and inequality has decreased between the two periods. Further if the percentage change on median is positive (i.e. there is growth) when $\mathrm{d}k > 0$ then necessarily $\mathrm{d}y_m$ is positive (cf. the expression $\Delta \ln y^M$ in Appendix A.7). The combined impact of an increase in k and y_m on the median income of the poor will be positive and greater than for the median income (see Appendix). Thus, the relative change on the median income of the poor is higher than the relative change in the median income of the distribution. We show that the relative change on the mean income of the poor is also higher than the relative change on the mean income of the population. Thus, for both distributions, the comparison between the relative change in median and mean will shed light on the pro-poor nature (or not) of the policy and its inequality reducing effect.

[17]For recent developments of this literature, see Lambert (2010).

6 Conclusions

Taking two right skewed distributions (lognormal and Pareto), we have shown that the true theoretical value of the 'elasticity' of the income of the poor with respect to the income of the population is equal to 1. However, a regression of the mean income of the poor on the mean income of the population would produce a biased estimate lower than 1 if changes in inequality are present. The bias disappears when a control for inequality as an estimate of the shape of the distribution is introduced in the regression. However, as the underlying distribution of incomes is unknown, we have no method to accurately estimate this shape parameter.

We have proved that the measure based on the median income is robust to changes in inequality and provides an accurate estimate of the 'elasticity' of the income of the poor to the aggregate income, even for a random mix of distributions. In this sense, a structural shock or an economic policy that affects positively the median income of the distribution tends to have a proportional change on the median income of the poor, whatever the change in inequality. A policy whose impact is positive on the mean income can have any effect on the income of the poor depending on the accompanying change in inequality.

Pro-poor policies can be identified using the comparison between the percentage change in mean income and that in the median income as discussed in Section 5. A larger impact on the log of median income than on the log of mean income will tend to indicate a pro-poor policy. It will reduce relative poverty through a higher impact on incomes of the poor which also decreases the ratio of mean to median incomes.

Tables and Figures

Fig. 1: Dispersion terms of log-mean incomes for selected distributions

(a) Lognormal (b) Pareto

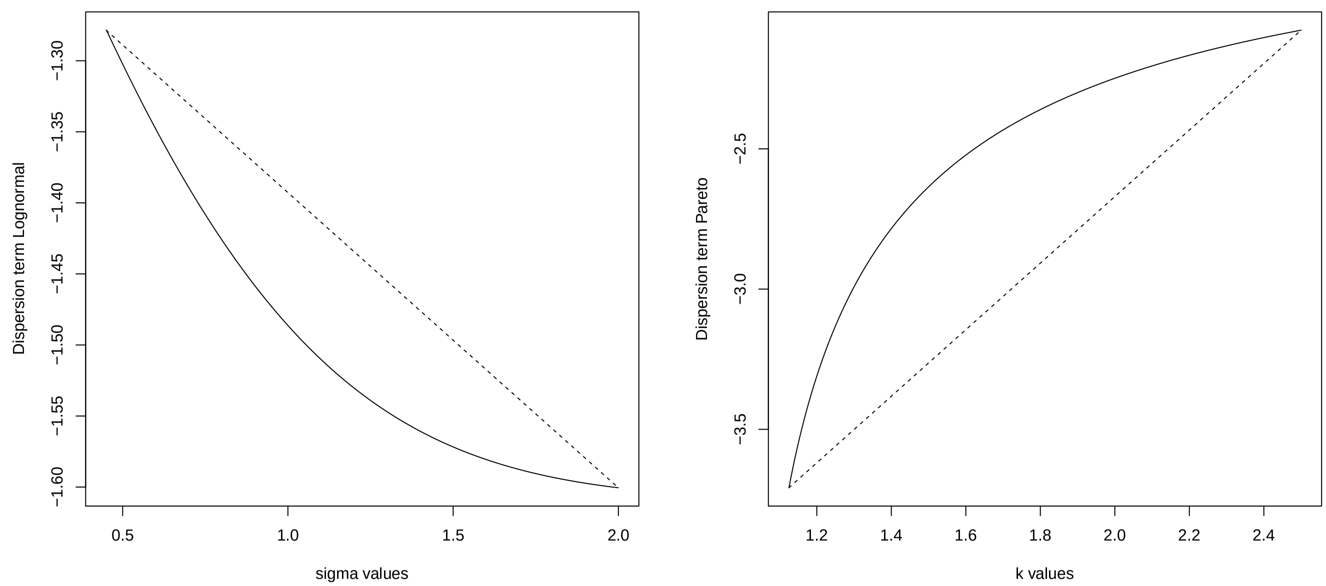


Fig. 2: Elasticity estimates for different distributions

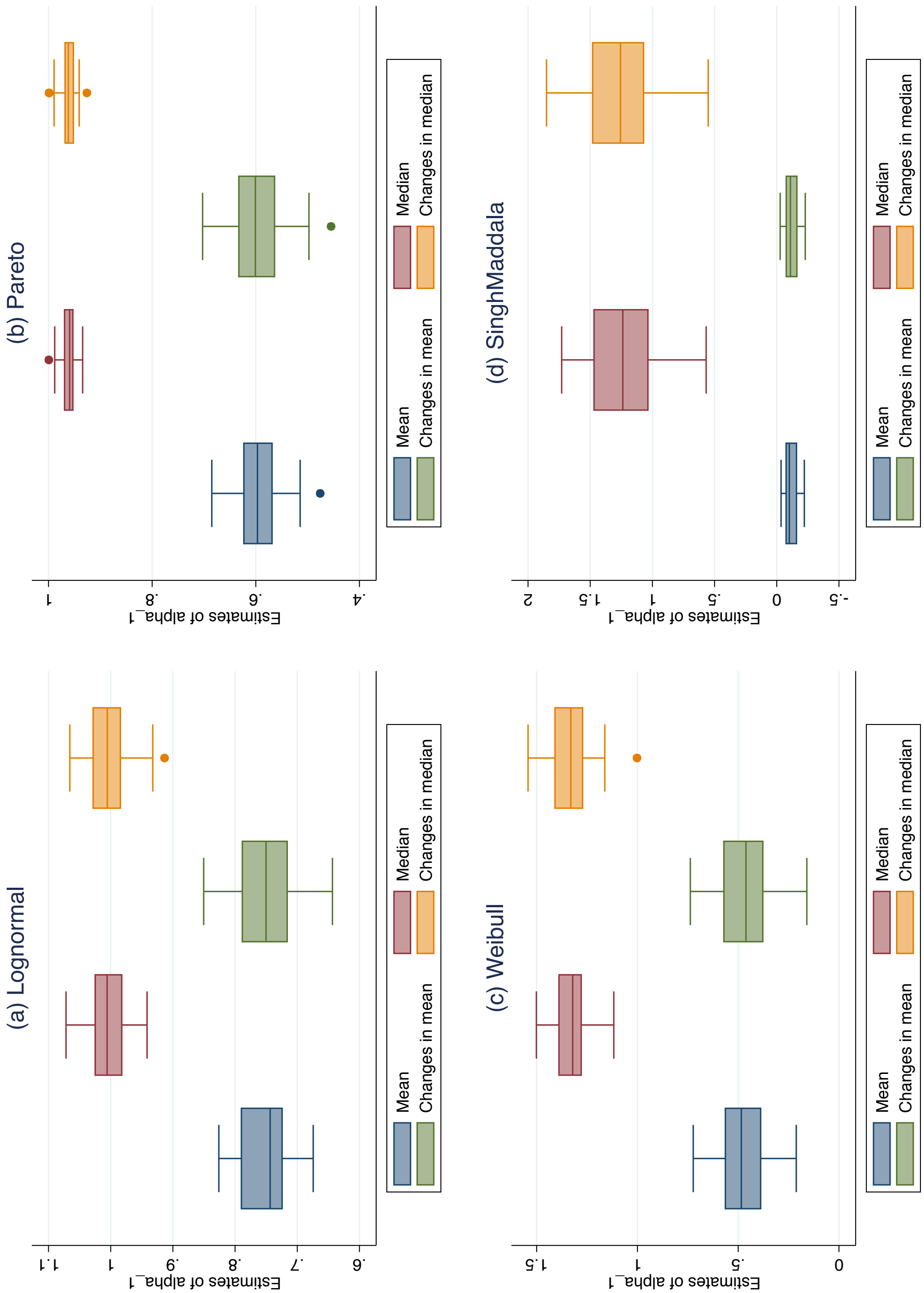
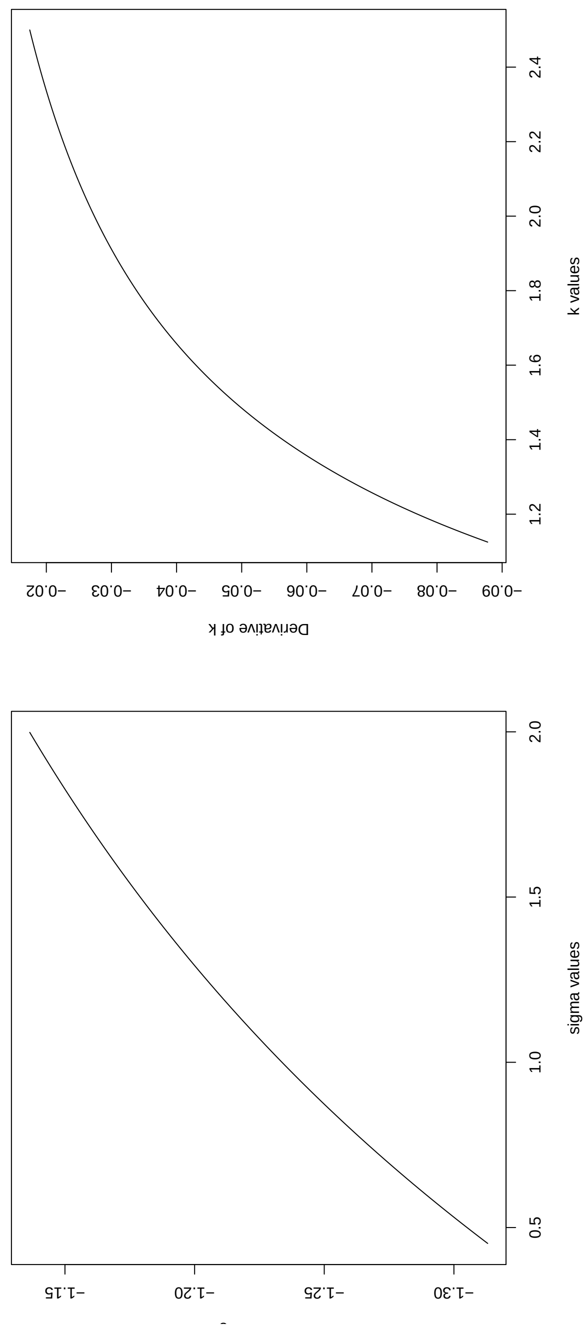


Fig. 3: Partial derivatives of log-mean income to changes in dispersion

(a) Lognormal (b) Pareto



Tab. 1: Summary of bias in $\hat{\alpha}_1$

	Lognormal	Pareto
Mean incomes	Biased $\hat{\alpha}_1 < 1$	Biased $\hat{\alpha}_1 < 1$
Median incomes	Unbiased $\hat{\alpha}_1 = 1$	Small bias $\hat{\alpha}_1 \approx 0.963$

Tab. 2: Regression of the measure of income of poor on the measure of income of the population

Lognormal distribution

	mean	median	Δmean	Δmedian
(Intercept)	0.3734	-1.4558^{**}	-0.0014	0.0024
	(0.4529)	(0.2768)	(0.1274)	(0.0850)
$\ln(m_D)$	0.4670^{**}	0.9633^{**}	0.2872^{**}	0.9293^{**}
	(0.0820)	(0.0599)	(0.0862)	(0.0745)
Adjusted R^2	0.2411	0.7220	0.0935	0.6122

Pareto distribution

	mean	median	Δmean	Δmedian
(Intercept)	2.6109^{**}	-0.3074^{*}	-0.0054	-0.0000
	(0.5055)	(0.1491)	(0.0468)	(0.0124)
$\ln(m_D)$	0.6056^{**}	0.9950^{**}	0.5504^{**}	0.9957^{**}
	(0.0568)	(0.0179)	(0.0602)	(0.0020)
Adjusted R^2	0.5327	0.9691	0.4571	0.9618

**, * statistically significant at 1 % and 5 % level with respect to zero. In parenthesis, the standard deviation of estimates.

Tab. 3: Regressions with US income data

	mean	median
(Intercept)	0.378**	-0.148*
	(0.0817)	(0.0609)
$\ln(m_D)$	0.825**	0.892**
	(0.00785)	(0.00596)
Observations	44	44
Adjusted R^2	0.996	0.998

**, * statistically significant at 1 % and 5 % level with respect to zero. In parenthesis, the standard deviation of estimates.

Tab. 4: Regressions in a panel of countries with income shares (1978-2012)

	1978-2012		2005-2012		last observation	
	mean	median	mean	median	mean	median
(Intercept)	-1.048***	-0.914***	-0.824***	-0.767***	-0.803***	-0.763***
	(0.112)	(0.070)	(0.190)	(0.121)	(0.229)	(0.164)
$\ln(m_D)$	0.967***	0.994***	0.947***	0.980***	0.948***	0.972***
	(0.015)	(0.010)	(0.024)	(0.016)	(0.031)	(0.023)
F(alpha1 = 1)	5.273	0.455	4.709	1.461	2.858	1.526
p-value	0.0219	0.500	0.0309	0.228	0.0942	0.220
Observations	859	859	266	266	98	98
R^2	0.838	0.927	0.850	0.933	0.909	0.950

**, * statistically significant at 1 % and 5 % level. In parenthesis, the robust standard deviation of estimates.

Tab. 5: Changes in poor's income following reductions in inequality

	Lognormal $\sigma \downarrow$	Pareto $k \uparrow$
Mean income of poor	Increase	Decrease
Median income of the poor	Increase	Decrease

A Appendices

A.1 Calculations of the mean and median income for the bottom quintile

The mean income over a defined interval is the conditional expectation of the distribution on this interval. The expression of this expectation is :

$$\overline{y}_p = \int_{Q1} y f_{Q1}(y \mid y \in Q1) dy,$$

where $f_{Q1}(\cdot)$ is the conditional density function for the interval $Q1$, the bottom quintile of the distribution. Here, the interval $Q1$ is defined by $[0^+, t_{0.2}]$, where $t_{0.2}$ is the 20% percentile of the distribution. Thus the conditional density function $f_{Q1}(\cdot)$ can be obtained from the probability distribution function $F_{Q1}(\cdot)$:

$$F_{Q1}(y \mid y \in Q1) = \frac{F(y)}{F(t_{0.2})} = 5 \times F(y)$$
$$\Rightarrow f_{Q1}(y \mid y \in Q1) = 5 \times f(y) \qquad \text{and} \qquad \overline{y}_p = 5 \times \int_{Q1} y f(y) dy.$$

<u>*Lognormal distribution*</u>

The underlying distribution is a Normal distribution with the same parameters and we can replace $t_{0.2}^{LN}$ by the monotonic transformation applied to the bottom quintile in the Normal distribution, $\exp(\mu + z_{0.2}\sigma)$ where $z_{0.2}$ is the upper bound of the bottom quintile of the Normal distribution $N(0,1)$. Using the definition of the mean of a truncated lognormal distribution, we obtain :

$$\begin{aligned}\overline{y}_p^{LN} &= 5 \times \left(\int_{0^+}^{\infty} y f_{\mu,\sigma}(y) dy - \int_{\overline{Q1}} y f_{\mu,\sigma}(y) dy \right) \\ &= 5 \times \left(\exp\left(\mu + \frac{\sigma^2}{2}\right) - \exp\left(\mu + \frac{\sigma^2}{2}\right) \Phi\left[\frac{-\ln(k) + \mu + \sigma^2}{\sigma}\right] \right),\end{aligned}$$

where $\overline{Q1}$ is the interval complementary to the bottom quintile, Φ the cumulative distribution of a Normal $N(0,1)$ and k is the lower bound of the interval $\overline{Q1}$. Finally, replacing k we have :

$$\overline{y}_p^{LN} = 5 \times \exp\left(\mu + \frac{\sigma^2}{2}\right) \times [1 - \Phi(-z_{0.2} + \sigma)]$$

The median income of the poor divides the bottom quintile into two non-overlapping sub-samples. Each sub-sample contains the same number (percentage) of observations. It is the definition of the 10 % percentile of the distribution $f(\cdot)$. With the results obtained before, the median income of the poor under log-normality can be derived as follows :

$$y_p^{M,LN} = t_{0.1}^{LN} = \exp(\mu + z_{0.1}\sigma)$$

Pareto distribution

The density function is given by :

$$f(y) = \frac{k y_m^k}{y^{k+1}} \qquad \text{with} \quad y_m > 0 \quad \text{and} \quad k > 1$$

where y_m is the lower possible value of the distribution and the restriction over the parameter of dispersion k is necessary for the existence of the mean value of the distribution (income per capita). The bottom quintile for a Pareto distribution $P(y_m, k)$ is delimited by $\left[y_m, .y_m \times (0.8)^{-\frac{1}{k}}\right]$. The mean income for the bottom quintile is the conditional expectation over this interval and the median income of the poor corresponds to the 10th percentile of the distribution :

$$\overline{y}_p^{P} = 5 \times \frac{k y_m}{k-1} \times \left[1 - 0.8^{1-\frac{1}{k}}\right]$$
$$y_p^{M,P} = t_{0.1}^{P} = y_m \times 0.9^{-\frac{1}{k}}$$

A.2 Relationship between the Gini index and the dispersion parameters of the distributions (σ and k)

The Gini index is a function of the dispersion of the underlying distribution. So, the dispersion parameter can be obtained once the Gini index is known. When the distribution is assumed to be a lognormal distribution, dispersion is given by the following formula [18]:

$$\sigma = \sqrt{2}\ \Phi^{-1}\left(\frac{1+G}{2}\right)$$

where Φ^{-1} is the inverse cumulative distribution function of a normal distribution $N(0,1)$ and G is the Gini index, $G \in [0,1]$.

In the case of the Pareto distribution, the parameter k is determined by :

$$k = \frac{1}{2} + \frac{1}{2G}.$$

A.3 Covariance calculations

In what follows, as an illustration, we work out $Cov\left(\frac{\sigma^2}{2};\sigma\right)$ and $Cov(\ln\left(\frac{k}{k-1}\right);k)$ for $\sigma \sim Uniform[a,b]$ and $k \sim Uniform[a,b]$.

$$Cov\left(\frac{\sigma^2}{2};\sigma\right) = \frac{1}{2} \times Cov(\sigma^2;\sigma)$$

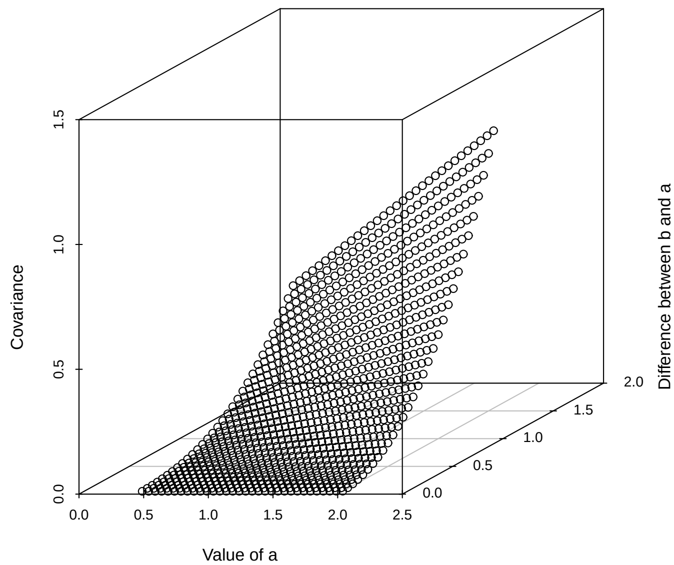
$$\begin{aligned}
Cov(\sigma^2;\sigma) &= E[\sigma^3] - E[\sigma^2]E[\sigma]. \\
E[\sigma^3] &= \int_a^b \sigma^3 \times \frac{1}{b-a} d\sigma = \left.\frac{\sigma^4}{4\times(b-a)}\right|_a^b = \frac{b^4-a^4}{4\times(b-a)}. \\
E[\sigma^2] &= \int_a^b \sigma^2 \times \frac{1}{b-a} d\sigma = \left.\frac{\sigma^3}{3\times(b-a)}\right|_a^b = \frac{b^3-a^3}{3\times(b-a)}.
\end{aligned}$$

$$\begin{aligned}
Cov(\sigma^2;\sigma) &= \frac{b^4-a^4}{4\times(b-a)} - \frac{b^3-a^3}{3\times(b-a)} \times \frac{a+b}{2} \\
&= \frac{1}{12\times(b-a)} \times \left[3b^4 - 3a^4 - 2ab^3 - 2b^4 + 2a^4 + 2a^3b\right] \\
&= \frac{1}{12\times(b-a)} \times \left[(b^4-a^4) - 2\times(ab^3 - a^3b)\right] \\
&= \frac{1}{12\times(b-a)} \times \left[(b^2+a^2)\times(b^2-a^2) - 2ab\times(b^2-a^2)\right] \\
&= \frac{b+a}{12} \times \left[b^2 - 2ab + a^2\right] = \frac{b+a}{12} \times (b-a)^2.
\end{aligned}$$

For $[a,b] \in \mathbb{R}^2_{++}$, $Cov(\sigma^2;\sigma) > 0$.

[18] See Aitchinson and Brown (1966).

Fig. A.1: Covariance between σ and σ^2



Now, let us turn to $Cov\left(\ln\left(\frac{k}{k-1}\right);k\right)$.

We calculate the cumulative density function of $z=\ln\left(\frac{k}{k-1}\right)$:

$$
\begin{aligned}
G(Z<z) &= P\left[\ln\left(1+\frac{1}{k-1}\right)<z\right] = P\left[\frac{1}{k-1}<e^z-1\right] \\
&= P\left[\frac{e^z}{e^z-1}<k\right] = 1-\left.\frac{k-a}{b-a}\right|_a^{\frac{e^z}{e^z-1}} = \frac{1}{b-a}\times\left(b-a-\frac{e^z}{e^z-1}+a\right) \\
&= \frac{1}{b-a}\times\left(b-1-\frac{1}{e^z-1}\right). \\
g(z) &= \frac{1}{(b-a)(e^z-1)^2}\times e^z.
\end{aligned}
$$

For $1 < a < b$, we have that $\frac{a}{a-1} > \frac{b}{b-1}$, then $a^* = \ln\left(\frac{a}{a-1}\right) > b^* = \ln\left(\frac{b}{b-1}\right)$.

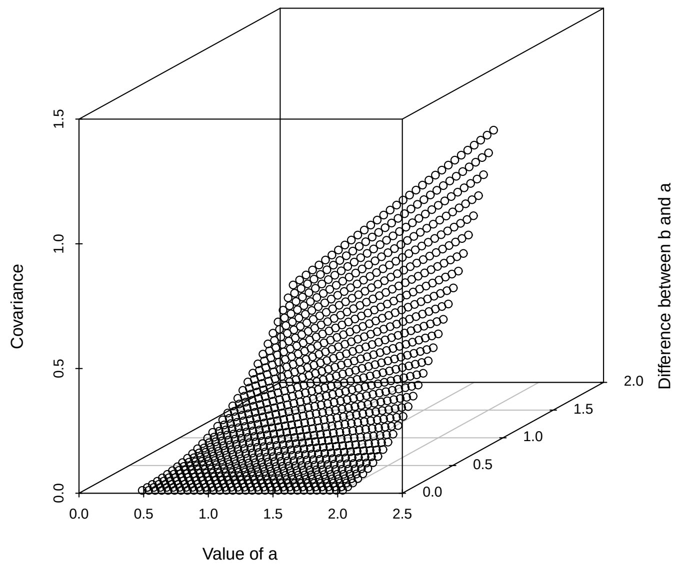
$$\begin{aligned}
E[Z] &= \int_{b^*}^{a^*} \frac{ze^z}{(b-a)\times(e^z-1)^2}dz = \frac{1}{b-a}\times\left[-\left.\frac{z}{e^z-1}\right|_{b^*}^{a^*} + \int_{b^*}^{a^*}\frac{1}{e^z-1}dz\right]\\
&= \frac{1}{b-a}\times\left.\left(-\frac{z}{e^z-1}+\ln(e^z-1)-z\right)\right|_{b^*}^{a^*} =\\
&= \frac{1}{b-a}\times\left[-(a-1)\ln\left(\frac{a}{a-1}\right)-\ln(a-1)-\ln\left(\frac{a}{a-1}\right)\right.\\
&\qquad\left.+(b-1)\ln\left(\frac{b}{b-1}\right)+\ln(b-1)+\ln\left(\frac{b}{b-1}\right)\right]\\
&= \frac{1}{b-a}\times\left[b\ln\left(\frac{b}{b-1}\right)-a\ln\left(\frac{a}{a-1}\right)+\ln(b-1)-\ln(a-1)\right]\\
&= \frac{1}{b-a}\times\left[b\ln(b)-a\ln(a)+(1-b)\ln(b-1)-(1-a)\ln(a-1)\right].
\end{aligned}$$

$$\begin{aligned}
E[Zk] &= \int_a^b \frac{k}{b-a}\times\ln\left(\frac{k}{k-1}\right)dk = \frac{1}{b-a}\times\left[\ln\left(\frac{k}{k-1}\right)\times\left.\frac{k^2}{2}\right|_a^b+\right.\\
&\qquad\left.+\int_a^b\frac{k-1}{k}\times\frac{-1}{(k-1)^2}\times\frac{k^2}{2}dk\right]\\
&= \frac{1}{2(b-a)}\times\left.\left[\ln\left(\frac{k}{k-1}\right)\times k^2+k+\ln(k-1)\right]\right|_a^b\\
&= \frac{1}{2(b-a)}\times\left[b^2\ln(b)-a^2\ln(a)+b-a+(1-b^2)\ln(b-1)-(1-a^2)\ln(a-1)\right].
\end{aligned}$$

$$\begin{aligned}
Cov(\ln\left(\frac{k}{k-1}\right);k) &= \frac{1}{2(b-a)}\times\left[b^2\ln(b)-a^2\ln(a)+b-a+(1-b^2)\ln(b-1)\right]\\
&\qquad-(1-a^2)\ln(a-1)-\frac{a+b}{2}\times\frac{1}{b-a}\times[b\ln(b)-a\ln(a)\\
&\qquad+(1-b)\ln(b-1)-(1-a)\ln(a-1)]\\
&= \frac{1}{2(b-a)}\times[ab\ln(a)-ab\ln(b)+(b-a)\\
&\qquad+(1-a-b+ab)\ln(b-1)-(1-a-b+ab)\ln(a-1)]\\
&= \frac{1}{2(b-a)}\times[ab(\ln(a)-\ln(b))+(b-a)\\
&\qquad+(1-a-b+ab)(\ln(b-1)-\ln(a-1))].
\end{aligned}$$

For $[a,b]\in\mathbb{R}^2_{++}$ where $b>a$, $Cov\left(\ln\left(\frac{k}{k-1}\right);k\right)<0$.

Fig. A.2: Covariance between k and $\ln\left(\frac{k}{k-1}\right)$



A.4 Theoretical estimate of $\hat{\alpha}_1$ when using median incomes and the Pareto distribution

The logarithm of the median income of the poor is :

$$\begin{aligned}
\ln y_p^{M,P} &= \ln y_m - \frac{1}{k}\ln 0.9 \\
&= \ln y_m - \frac{1}{k}\ln(0.5) - \frac{1}{k}\ln\left(\frac{9}{5}\right) \\
&= \ln y^{M,P} - \frac{1}{k}\ln\left(\frac{9}{5}\right).
\end{aligned}$$

Replacing this relationship in the expression of the estimate of the elasticity when median incomes of the Pareto distribution are used:

$$
\begin{aligned}
\hat{\alpha}_1^{M,P} &= \frac{Cov(\ln y_p^{M,P}; \ln y^{M,P})}{V(\ln y^{M,P})} \\
&= \frac{E\left[\ln y_p^{M,P} \ln y^{M,P}\right] - E\left[\ln y_p^{M,P}\right] E\left[\ln y^{M,P}\right]}{V(\ln y^{M,P})} \\
&= \frac{V(\ln y^{M,P}) + E\left(\frac{\ln(0.5)\times\ln\left(\frac{9}{5}\right)}{k^2}\right) - E\left(\frac{\ln 0.5}{k}\right) E\left(\frac{\ln\left(\frac{9}{5}\right)}{k}\right)}{V(\ln y^{M,P})} \\
&= 1 + \ln(0.5) \times \ln\left(\frac{9}{5}\right) \times \frac{V\left(\frac{1}{k}\right)}{V(\ln y^{M,P})}
\end{aligned}
$$

where $V(\ln y^{M,P}) = V(\ln y_m) + (\ln 0.5)^2 \times V\left(\frac{1}{k}\right)$. We need to derive the variance of the logarithm of y_m, $Z = \ln y_m$, where $y_m \sim U[a,b]$ and the variance of the logarithm of $\frac{1}{k}$, $T = \frac{1}{k}$, where $k \sim U[c,d]$.

$$
\begin{aligned}
G(Z < z) &= P\left[\ln y_m < z\right] = P\left[y_m < e^z\right] \\
&= \left.\frac{y-a}{b-a}\right|_a^{e^z} = \frac{e^z - a}{b-a}. \\
g(z) &= \frac{e^z}{b-a}.
\end{aligned}
$$

$$
\begin{aligned}
E[Z] &= \int_{\ln a}^{\ln b} \frac{ze^z}{b-a} dz = \frac{1}{b-a}\left[ze^z|_{\ln a}^{\ln b} - \int_{\ln a}^{\ln b} e^z dz\right] \\
&= \frac{1}{b-a} \times \left[ze^z - e^z|_{\ln a}^{\ln b}\right] \\
&= \frac{1}{b-a} \times \left[a - b + b\ln b - a\ln a\right]. \\
E[Z^2] &= \int_{\ln a}^{\ln b} \frac{z^2 e^z}{b-a} dz = \frac{1}{b-a}\left[z^2 e^z|_{\ln a}^{\ln b} - 2\int_{\ln a}^{\ln b} ze^z dz\right] \\
&= \frac{1}{b-a} \times \left[z^2 e^z - 2ze^z + 2e^z|_{\ln a}^{\ln b}\right].
\end{aligned}
$$

$$
\begin{aligned}
G(T<t) &= P\left[\frac{1}{k}<t\right]P\left[\frac{1}{t}<k\right] \\
&= 1-\left.\frac{k-c}{d-c}\right|_{c}^{\frac{1}{t}} = -\frac{1}{d-c}\times\left[\frac{1}{t}-c\right]. \\
g(t) &= \frac{1}{d-c}\times\frac{1}{t^2}. \\
E[T] &= \int_{\frac{1}{d}}^{\frac{1}{c}}\frac{1}{d-c}\times\frac{1}{t}dt = \frac{1}{d-c}\ln t|_{\frac{1}{d}}^{\frac{1}{c}} \\
&= \frac{\ln\frac{d}{c}}{d-c}. \\
E[T^2] &= \int_{\frac{1}{d}}^{\frac{1}{c}}\frac{1}{d-c}dt = \frac{1}{d-c}\, t|_{\frac{1}{d}}^{\frac{1}{c}} \\
&= \frac{1}{cd}.
\end{aligned}
$$

Given the values of a, b, c and d used for the simulations of Pareto distributions, we can calculate the theoretical bias and the theoretical estimate :

$$
\begin{aligned}
\hat{\alpha}_1^{M,P} &= 1+\ln(0.5)\times\ln\left(\frac{9}{5}\right)\times\frac{V\left(\frac{1}{k}\right)}{V(\ln y^{M,P})} \\
&= 1-.0374147 = .9625853
\end{aligned}
$$

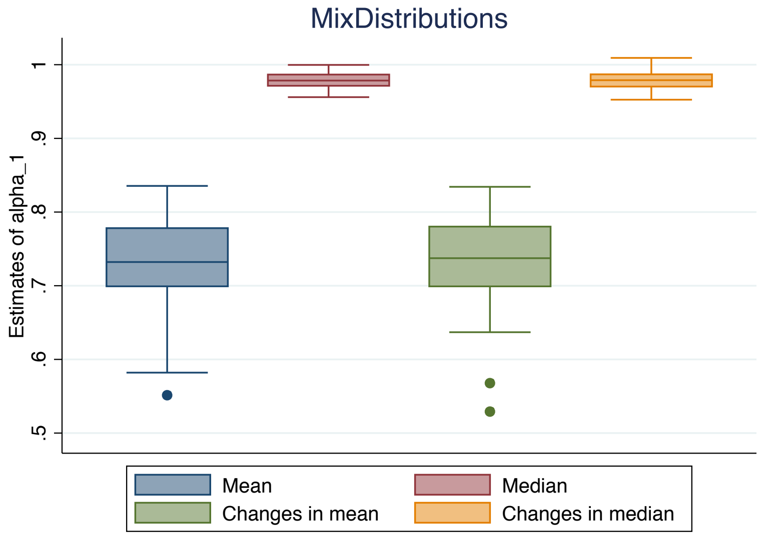
A.5 Simulations using a mix of distributions

A sample of 200 'mixes' of distributions is generated using the following procedure. Observations generated independently from a lognormal distribution and a Pareto distribution are mixed in 'random' proportions to generate a new 'mix' distribution[19]. The share of observations in the mix coming from each of the two distributions ranges from 25 to 75 % and is randomly chosen for every mixture. The unique condition imposed on the underlying distributions in each mixture is that they are defined over the same interval which implies that the minimum value is strictly positive. Once the mix of the two distributions is done, we compute the mean and median values for each mix which constitute one observation for our Equation 10. The procedure is repeated until the desired number of observations is achieved.

Figure A.3 below shows the results obtained by repeating the above experiment 40 times and estimating Equation 10 using the sample of 200 means and medians. It presents a boxplot for the estimate of α_1 in each version of the regression for the random mix

[19] Before computing every mix, new observations are generated for each distribution i.e. the lognormal and the Pareto distributions.

Fig. A.3: Elasticity estimates for lognormal and Pareto distributions



of distributions. It clearly shows that the elasticity in median incomes is larger than the elasticity in mean incomes and we find that the former elasticity is not statistically different from 1.

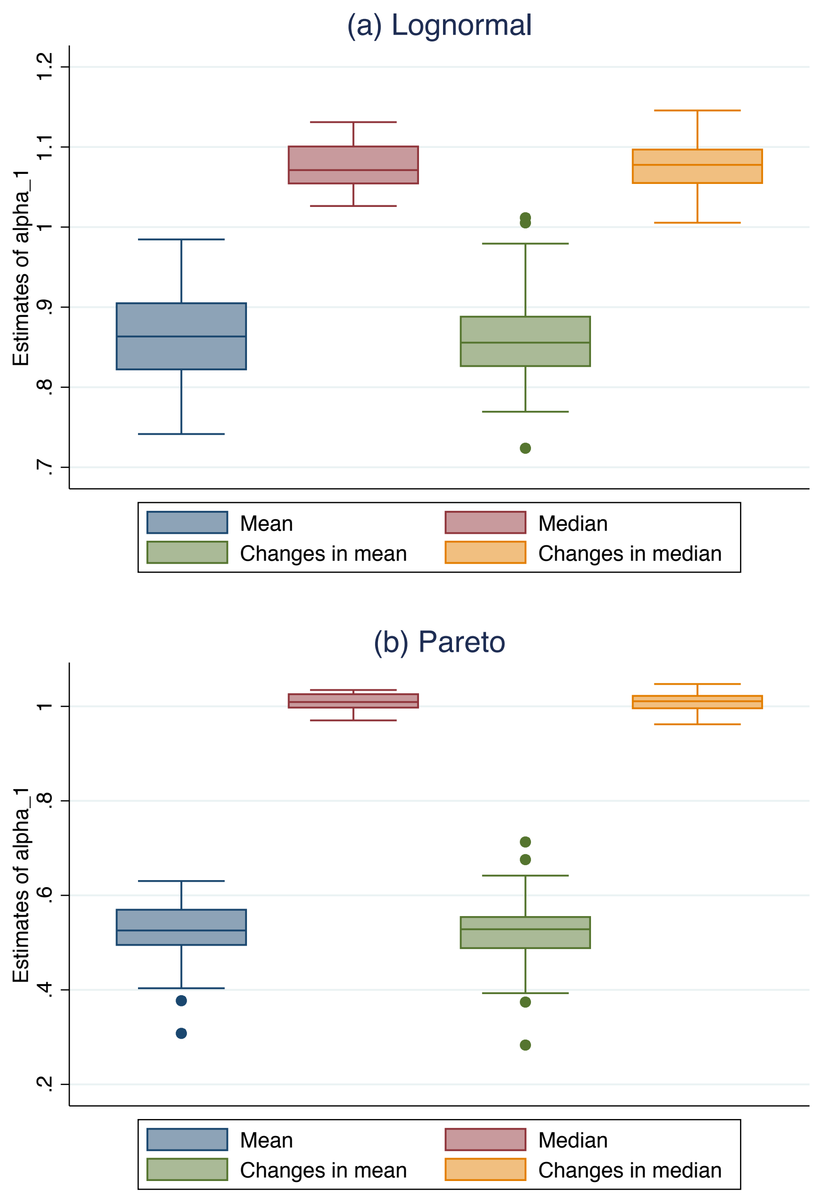
A.6 Estimates of α_1 under positive covariance between central tendency and dispersion in incomes

The assumption of no covariance ($Cov(\mu, \sigma) = Cov(\ln y_m, k) = 0$) imposed in section 3 was implemented in order to set a convenient benchmark for the estimates for α_1 and it could be seen as a restrictive analysis for the framework proposed by Equation 10. Here, we release this assumption for the Lognormal and Pareto distributions and run an analogous bootstrap analysis to the one presented in subsection 4.2.

On average, the increases in median incomes are associated with decreases in income inequality for both distributions but we cannot ensure this in every subsample of the bootstrap procedure due to the random generation of data. A doubling of median incomes is associated with a 5% decrease of the Gini index and we consider that this is a non-negligible characterization of 'pro-poor' growth spells.

Each of the 40 subsamples of the bootstrap procedure consists of 200 populations for which we compute the mean and median values required for Equation 10. Figure A.6 below shows the results obtained by releasing the assumption of no covariance. Under log-normality, the estimate based on median incomes is consistently larger than 1 and it captures the pro-poorness of growth described by our sample. On the other hand, the estimate based on mean incomes is still lower than 1 and it does not have the expected estimate due to the bias described in section 3. For the Pareto distribution, results are less clear for $\hat{\alpha}_1$ based on median incomes even though two thirds of the subsamples show an estimate larger than 1. Once again, the estimate based on mean incomes fails to capture the expected estimate.

Fig. A.4: Elasticity estimates for lognormal and Pareto distributions



A.7 Comparisons among changes in mean/median income of the population and changes in mean/median income of the poor

For the lognormal distribution :

Suppose we have

$$\begin{aligned}\Delta \ln y^{M,LN} &> \Delta \ln \overline{y}^{LN}\\ \mathrm{d}\mu &> \mathrm{d}\mu + \sigma \mathrm{d}\sigma\end{aligned}$$

This necessarily implies a reduction in inequality ($d\sigma < 0$) given that $\sigma > 0$. Furthermore, we have that :

$$\begin{aligned}\Delta \ln y_p^{M,LN} &= \mathrm{d}\mu + z_{0.1}\mathrm{d}\sigma > \mathrm{d}\mu = \Delta \ln y^{M,LN}\\ \Delta \ln \overline{y}_p^{LN} &= \mathrm{d}\mu + \left[\sigma - \frac{\phi(-z_{0.2}+\sigma)}{1-\Phi(-z_{0.2}+\sigma)}\right]\mathrm{d}\sigma > \mathrm{d}\mu > \Delta \ln \overline{y}^{LN}\end{aligned}$$

Thus

$$\begin{aligned}\Delta \ln y_p^{M,LN} &> \Delta \ln y^{M,LN}\\ \Delta \ln \overline{y}_p^{LN} &> \Delta \ln \overline{y}^{LN}\end{aligned}$$

For the Pareto distribution :

Suppose

$$\begin{aligned}\Delta \ln y^{M,P} &> \Delta \ln \overline{y}^{P}\\ \frac{\mathrm{d}y_m}{y_m} + \frac{\ln(0.5)\mathrm{d}k}{k^2} &> \frac{\mathrm{d}y_m}{y_m} - \frac{\mathrm{d}k}{k(k-1)}\end{aligned}$$

Given that $\frac{\ln(0.5)}{k^2} > \frac{\ln(0.5)}{k^2-k} > -\frac{1}{k(k-1)}$ for $k > 1$, the above inequality implies $\mathrm{d}k > 0$ i.e. a reduction of inequality. Further, $\mathrm{d}k > 0$ and $\Delta \ln y^{M,P} > 0$ imply $\mathrm{d}y_m > 0$.

Thus, for the bottom quintile of the distribution, we have :

$$\begin{aligned}\Delta \ln y_p^{M,P} &= \frac{\mathrm{d}y_m}{y_m} + \frac{\ln(0.9)\mathrm{d}k}{k^2} > \frac{\mathrm{d}y_m}{y_m} + \frac{\ln(0.5)\mathrm{d}k}{k^2} > 0\\ \Delta \ln \overline{y}_p^{P} &= \frac{\mathrm{d}y_m}{y_m} + \left[-\frac{1}{k(k-1)} - \frac{0.8^{1-\frac{1}{k}}}{1-0.8^{1-\frac{1}{k}}} \times \frac{\ln(0.8)}{k^2}\right]\mathrm{d}k\\ &> \frac{\mathrm{d}y_m}{y_m} - 0.1 \times \mathrm{d}k > \frac{\mathrm{d}y_m}{y_m} + \frac{\ln(0.5)\mathrm{d}k}{k^2} > 0\end{aligned}$$

the before last inequality being valid for our range of k values (approximately as long as $k < \sqrt{\frac{\ln 0.5}{-0.1}} \approx 2.63$).

Therefore,

$$\Delta \ln y_p^{M,P} > \Delta \ln y^{M,P},$$
$$\Delta \ln \overline{y}_p^{P} > \Delta \ln \overline{y}^{P}.$$

References

[1] Adams, Richard H. , 2004. Economic growth, inequality and poverty: Estimating the growth elasticity of poverty. World Development, 32(12), pp. 1989-2014.

[2] Aitchison, J., and J. Brown, 1966. The Log-Normal Distribution. Cambridge University Press.

[3] Bourguignon, Francois, 2004a. The growth elasticity of poverty reduction: Explaining heterogeneity across countries and time periods. World Bank Working Paper.

[4] Bourguignon, Francois, 2004b. The poverty-growth-inequality triangle. Indian Council for Research on International Economic Relations, New Delhi Working Papers 125, Indian Council for Research on International Economic Relations, New Delhi, India, March.

[5] Chen, Shaohua, and Martin Ravallion, 2001. How did the world's poorest fare in the 1990s ? Review of Income and Wealth, 47(3), pp. 283-300, September.

[6] Dollar, David, and Aart Kraay, 2002. Growth is good for the poor. Journal of Economic Growth, 7(3), pp. 195-225, September.

[7] Dollar, David, and Aart Kraay, 2004. Trade, growth and poverty. The Economic Journal, 114, pp. F22-F49.

[8] Duclos, Jean-Yves, 2009. What is 'pro-poor' ? Social Choice and Welfare, 32(1), pp.37-58, January.

[9] Gourdon, Julien, Nicolas Maystre, and Jaime de Melo, 2008. Openness, inequality and poverty: Endowments matter. The Journal of International Trade and Economic Development, 17(3), pp. 343-378.

[10] Heltberg, Rasmus, 2002. The poverty elasticity of growth. WIDER Discussion Paper 2002/21, UNU-WIDER.

[11] Kakwani, Nanak, 1993. Poverty and economic growth with application to Cote d'Ivoire. Review of Income and Wealth, 39(2), pp. 121-39, June.

[12] Kakwani, Nanak, and Ernesto Pernia, 2000. What is pro-poor growth ? Asian Development Review, 18(1), pp. 1-16.

[13] Klasen, Stephan, 2008. Economic growth and poverty reduction: Measurement issues using income and non-income indicators. World Development, 36(3), pp. 420-445.

[14] Klasen, Stephan, and Mark Misselhorn, 2008. Determinants of the growth semi-elasticity of poverty reduction. Ibero America Institute for Econ. Research (IAI) Discussion Papers 176, Ibero-America Institute for Economic Research.

[15] Kraay, Aart, 2006. When is growth pro-poor ? Evidence from a panel of countries. Journal of Development Economics, 80(1), pp. 198-227, June.

[16] Lambert, Peter J. , 2010. Pro-poor growth and the lognormal distribution. Journal of Income Distribution, 19(3-4), pp. 88-99, September.

[17] Lombardo, Vincenzo, 2009. Growth and inequality effects on poverty reduction in Italy. MPRA Paper 14351, University Library of Munich, Germany, February.

[18] Lopez, Humberto, and Luis Serven, 2006. A normal relationship ? poverty, growth, and inequality. Policy Research Working Paper Series 3814, The World Bank, January.

[19] Ravallion, Martin, 2001. Growth, inequality and poverty: Looking beyond averages. World Development, 29(11), pp. 1803-1815, November.

[20] Ravallion, Martin, 2004. Pro-poor growth : A primer. Policy Research Working Paper Series 3242, The World Bank, March.

[21] Ravallion, Martin, 2005. A poverty-inequality trade off ? Journal of Economic Inequality, 3(2), pp. 169-181, August.

[22] Ravallion, Martin, and Shaohua Chen, 2003. Measuring pro-poor growth. Economics Letters, 78(1), pp. 93-99, January.

[23] Ravallion, Martin, and Gaurav Datt, 2002. Why has economic growth been more pro-poor in some states of India than others ? Journal of Development Economics, 68(2), pp. 381-400, August.

[24] Ravallion, Martin, and Monika Huppi, 1991. Measuring changes in poverty: Amethodological case study of Indonesia during an adjustment period. World Bank Economic Review, 5(1), pp. 57-82, January.

[25] Sala-i Martin, Xavier, 2006. The world distribution of income: Falling poverty and ... convergence, period. The Quarterly Journal of Economics, 121(2), pp. 351-397, May.

[26] Son, Hyun Hwa, 2004 . A note on pro-poor growth. Economics Letters, 82(3), pp. 307-314, March.

Chapter II

Weak Links and Growth

CHAPTER II
Weak Links and Growth

Cristian Ugarte[*]

Department of Economics

University of Geneva, Switzerland

Abstract

The presence of *Weak Links* can hurt economic growth as shown by Jones (2011). In this paper we test this proposition. We first build an empirical measure of *Weak Links* that identifies sectors with low productivity that tend to be non tradeable and that are heavily used as intermediate inputs by other sectors. We then estimate a growth regression to quantify the impact of *Weak Links* on growth. A 10 percent increase in the probability of observing a *Weak Link* sector leads to almost a 1 percentage point loss in annual growth. The lower is the productivity of *Weak Link* sectors, the lower is economic growth, and the more statistically significant becomes the impact of those *Weak Links* on economic growth.

Keywords: Growth, Development, Weak Links, Industrial Policy.

[*]I am grateful to Olivier Cadot, Celine Carrere, Pramila Crivelli, Jorge Davalos, Jaime de Melo, Ndiame Diop, Jonathan Eaton, Caroline Freund, Julien Gourdon, Marcelo Olarreaga, Frederic Robert-Nicoud, Silvio Tai and Laura Zoratto for very helpful discussions as well as all participants of the DSEC Informal Workshop at University of Geneva and the 4th Bolivian Conference on Development Economics. All remaining errors are mine. E-mail: cristian.ugarte@unige.ch

Contents

List of Tables

List of Figures

1 Introduction

In a recent paper Jones (2011) builds on two old ideas in development economics, i.e. linkages between activities and complementarities in chain productivity[1], to help explain the observed large differences in income levels across nations beyond the productivity differentials of primary factors such as capital and labor. Take the example of textile production. Knitting machines and (skilled) operators are essential to achieve production but it would not be possible to produce if electricity supply is nonexistent or its price is not affordable. Provision of electricity, a largely non tradeable input, and its cost of production will determine the profitability of the textile industry and hence the productivity that this sector might attain. Furthermore, electricity is not the only complementary good or service requested. For example, transport services will have a similar impact on sales' perspectives. Indeed, there is a long list of services or activities affecting productivity of other sectors. This sequence of activities is known as a production chain. As emphasized by Kremer (1993), several activities in this production chain might lower the productivity of a particular sector or even underpin the failure of that activity.

Thus, a complex list of intermediate goods and services, e.g electricity, fuel supply and transport services among others, has an impact on the productivity achieved by other sectors. This suggests that complementarity across production activities should be carefully considered if we seek to explain differentials in industrialization and development across countries. Jones (2011) integrates this rationale and shows that the aggregate output level of countries depends positively on a composite of complementarities. Sectors with low productivity will be the *Weak Link* of the aggregate of complementarities and consequently, lower levels of aggregate productivity will be associated with the presence of relatively low productive sectors in the economy.

The objective of this paper is twofold. First, we build a measure that helps identify the presence of *Weak Link* sectors using internationally comparable data for manufacture productivity for more than 100 countries covering the period 1963-2001. Sectors with low productivity tend to be non-tradeable and their substitution by similar products manufactured in the rest of the world is limited. *Weak Link* sectors are used as intermediate inputs by other sectors in the economy which makes them complementary to production in other sectors. Second, we quantify the impact on economic growth of the presence of *Weak Link* sectors in an economy using a simple growth regression framework.

Results suggest that a 10 percent increase in the probability of observing a *Weak Link* sector leads to almost a 1 percentage point loss in annual growth. The lower is the

[1]Hirschman (1958) was the first to consider the role played by linkages and complementarities. More recently, Kremer (1993) showed that small differences in quality and worker skills lead to large differences in output.

productivity of *Weak Link* sectors relative to the average productivity in the economy, the lower is economic growth and the more statistically significant becomes the impact. The impact of *Weak Links* on growth does not seem to vary with income levels as there are no statistically significant differences for developed and developing countries. Results are also robust to instrumenting for *Weak Links* using measures that do not depend on their tradeability or their use as an intermediate input.

The paper is organized as follows. The next section presents the most important features of Jones' model (2011). Section 3 presents the strategy adopted to detect the presence of low productivity in intermediate goods and summarizes some of its characteristics. Section 4 evaluates the impact of productivity deviations on annual growth rates in different setups. The last section summarizes the conclusions of the empirical analysis and gives some economic policy implications regarding industrialization and development.

2 Theoretical framework

Jones (2011) models an economy[2] where a continuum of goods is produced using the combination of a nested Cobb-Douglas function of physical capital K_i, human capital H_i and a composite of intermediate goods X_i used in the production of Q_i :

$$Q_i = A_i(K_i^{\alpha} H_i^{1-\alpha})^{1-\sigma} X_i^{\sigma}, \tag{1}$$

where Q_i is the total production of sector i and A_i is the total factor productivity of the sector.

Goods can be used as an intermediate inputs for the production of other goods or as final consumption goods ($Q_i = z_i + c_i$). The gross domestic production is the aggregation of all final uses as a single final good using a CES-aggregator for the continuum of goods with an elasticity of substitution $(1/1-\theta)$ greater than one $(0 < \theta < 1)$:

$$Y = \left(\int_0^1 c_i^{\theta} di \right)^{1/\theta} \tag{2}$$

Intermediate goods can be similarly aggregated but the elasticity of substitution between intermediates $(1/(1-\rho))$ is less than one $(-\infty < \rho < 0)$ reflecting the assumption that substitution is higher in final goods' consumption (across c_i) than in intermediate goods' consumption (across z_i). In fact, the possibility of substitutability exists but it significantly lower while replacing intermediate goods.

[2] A simplified version of Jones (2011) is used here as it ignores the presence of distortions at the sectoral level such as theft, labor restrictions or any form of expropriation which can be measured as an advalorem tax equivalent.

For a given level of human capital and physical capital, the expression of the domestic gross production can be solved for. Jones shows that in a competitive allocation of resources, the expression for GDP at the steady state is given by :

$$Y = \psi(\sigma)(B_\theta^{1-\sigma} B_\rho^{\sigma})^{\frac{1}{1-\sigma}} K^\alpha H^{1-\alpha} \tag{3}$$

$$\text{with} \quad B_\rho = \left(\int_0^1 (A_i)^{\frac{\rho}{1-\rho}} di \right)^{\frac{1-\rho}{\rho}} \quad \text{and} \quad \psi(\sigma) = (1-\sigma)\sigma^{\frac{\sigma}{1-\sigma}}.$$

In Equation 3, σ measures the importance of intermediate inputs in the production function, α is the capital's share of output if linkages between sectors do not exist ($\sigma = 0$) and B_θ is defined in a way analogous to B_ρ. The total factor productivity of the economy depends on the combination of these two composite terms of productivity and this is where complementarities and linkages play an important role. In fact, the degree of complementarity/substitution in final consumption (θ) and intermediate goods (ρ) will directly affect the aggregate composites of productivities in the previous expression. Furthermore, the effect of complementarities in intermediates goods increases for stronger linkages between sectors (σ) and it will significantly affect the total factor productivity of the economy for relatively low values of interaction between sectors ($\sigma \approx 0.1$[3]).

Given that both productivity composites, B_θ and B_ρ, are defined using the same aggregation function where only the value of the exponent changes for a given set of sectoral TFPs $\{A_i\}$, we can simultaneously analyzed their behavior. Indeed, the aggregator used for these productivity composites is known as a generalized mean which is an increasing function of the exponent. Given than values of θ are positive and thus the exponent of B_θ ranges between 0 and 1, the average calculated in B_θ is comprised between the geometric and the arithmetic means of sectoral productivities. On the other hand, given that ρ is negative and the values of the exponent of B_ρ ranges between -1 and 0 which implies that B_ρ ranges between the harmonic and the geometric means of sectoral productivities. In both cases, higher substituability across sectors (higher values of θ or ρ) leads to a higher value of the productivity composites due the monotonocity of the aggregator. For further details on the behavior of these aggregate, see Appendix A.1.

In the extreme case where substitution of intermediate and final goods are the lowest ($\rho \to -\infty$, $\theta \to 0^+$), the TFP of the economy is the weighted product of the geometric and the harmonic means of productivities[4]. Indeed, the low productivities across sectors

[3]The TFP of the country is a geometric average of B_ρ and B_θ (B_ρ is always larger than B_θ) and their corresponding weights for the geometric average are σ and $(1-\sigma)$. For given values of B_ρ and B_θ, the higher the value of σ, the lower the TFP of the country and it can be shown that for low values of σ, the average would reduce substantially with respect to B_ρ.

[4]Figure A.1 in the Appendix illustrates all possible combinations of composites B_ρ and B_θ for a uniform distribution of productivities in range [1, 20].

are compensated by allocating more resources into those sectors in a competitive equilibrium[5] and therefore the TFP of the country is not driven by the lowest productivity across sectors. In the most favorable case (the highest substitutions in consumption and in production), the TFP of the economy is the weighted product of the arithmetic mean ($\theta \to 1/2$) and the geometric mean of sectoral productivities ($\rho \to 0^-$). Given the monotonic relationship between these average productivities (arithmetic mean > geometric mean > harmonic mean), the higher the substituability of sectors in intermediates and/or final goods, the higher the level of gross domestic production.

The use of intermediate goods in production (σ) is crucial since it determines the existence and the magnitude of the multiplier of productivities in intermediate goods. In case $\sigma = 0$, the previous framework would be simplified to a more usual version of the expression of the GDP where the total factor productivity of the economy is the average value of productivities across sectors[6]. However, several studies[7] have estimated the share of intermediate inputs in total inputs to be around 1/2 regardless of the level of development of countries and this is the reason why complementarity in intermediates should not be neglected.

The log-linearized specification of Equation 3 is given by :

$$\ln Y = \ln \psi(\sigma) + B_\theta + \frac{\sigma}{1-\sigma} \ln B_\rho + \alpha \ln K + (1-\alpha) \ln H \tag{4}$$

Clearly, θ and ρ are major determinants of the level of development as specific values of these parameters would lead to different productivity composites (B_θ and B_ρ) across sectors[8]. Ideally, we would like to calculate the terms B_θ and B_ρ in order to evaluate their contribution to the level of development of countries but we lack of estimations of these parameters across sectors at the country level[9]. Furthermore, given that the proxies for total factor productivities are residuals from a regression, positive and negative values for residual productivities but the generalized average described in terms B_θ and B_ρ are only defined for positive values.

Therefore, the objective of this paper is twofold. Firstly, to propose a measure that can be used as a proxy for the productivity aggregates in Equation 4, particularly for the term B_ρ. We propose to assess the likelihood of observing relatively low productivities in intermediate inputs which takes into account the mix of intermediate goods used

[5] Jones (2011) shows that this compensation can be decomposed in two effects, the substitution and the complementarity effects which have opposite signs.

[6] This assumes infinite substitutability in consumption.

[7] See Basu (1995) and Ciccone (2002).

[8] Appendix 1 does an analysis of the impact on productivity at the national level for different values of θ and ρ as well as for different distributions of productivities. The two most extreme cases were presented in this section.

[9] Imposing the same degree of substitution across all sectors in a given economy would not be suitable to our case since we seek to measure the easiness/difficulty to replace some particular sectors.

by the production structure of each country and the tradeability/substitutability of such intermediate goods. Secondly, we use this measure to estimate the impact of low productivity in intermediate goods on aggregate growth by using a standard growth regression framework. Next section presents the variable used in the empirical analysis.

3 Weak Links: measurement and characterization

Our objective is that the proxy proposed here focuses on measuring relative productivity and its contagion at the sectoral level by country. Therefore, we use data for 28[10] sectors provided at the 3-digit level of the International Standard Industrial Classification of All Economic Activities (ISIC) Revision 2 by the United Nations Industrial Development Organization (UNIDO) Industrial Statistics Database (INDSTAT2)[11]. This harmonized source of data for the period 1963-2001[12] allows us to compare data across countries and time periods however its coverage is limited to manufacture production[13].

To measure *Weak Links*, we will proceed in three steps. First, we will use a regression of labor productivity to approximate the total factor productivities in productivity composites B_θ and B_ρ described in Equation 3. In a second step, we estimate a kernel density for the productivity distribution observed in country c in year t. The aim of this kernel density estimation is to allow us to measure abnormal deviations to the left of the productivity distribution. In a final step, we readjust the kernel density of the previous step according to sectoral share of intermediate sales and an index of sectoral tradeability to ensure that the proposed measure reflects the country's needs in intermediate inputs and the easiness to replace domestically produced goods by imports.

Step 1: Generate proxy values for total factor productivities

We use a measure of labor productivity at the sectoral level namely the ratio of value-added and the number of workers in sector s in country c in year t, named $q_{c,s,t}$ hereafter, based on data available in the INDSTAT2 database. To avoid substantial differences between labor productivities and total factor productivities, we will not work with labor productivity directly, but with residuals of a regression of labor productivity ($\widehat{\mathrm{r.q}}_{c,s,t}$) on country-year ($\lambda_{c,t}$) and sector-year fixed effects (γ_{s_t}) as shown in Equations 5-7. The fixed effects included in Equation 6 allow us to control for several features of productivities among which the fact that more labor abundant countries will produce more labor intensive products and therefore we observe lower productivities throughout

[10]The number of sectors may vary per country and year due to misreporting in the data.

[11]http://www.unido.org/.

[12]This data was the only source with an extensive coverage of countries and time periods available until recently. An updated version of this dataset was made available in 2013 by UNIDO but it reduces significantly the number of sectors reported since it uses the ISIC Revision 3 (18 sectors instead of 28 in ISIC Revision 2).

[13]Table A.1 lists countries included in this study.

or the bias due to price shocks specific to some sectors or demand shocks above the business cycle. The following system of equations presents the procedure in detail :

$$\text{Labor productivity}_{c,s,t} = \mathrm{q}_{c,s,t} = \frac{\text{Value added}_{c,s,t}}{\text{Labor}_{c,s,t}} \tag{5}$$

$$\mathrm{q}_{c,s,t} = \sum_{c,t} \lambda_{c,t} + \sum_{s,t} \gamma_{s,t} + \varepsilon_{c,s,t} \tag{6}$$

$$\widehat{\mathrm{r.q}}_{c,s,t} = \widehat{\varepsilon}_{c,s,t} \tag{7}$$

The minimum level of residual productivity across sectors in a country, or even the minimum level of residual productivity in the economy across time are some of the potential indicators of *Weak Links* that could be considered. However, minimum levels will capture an extreme case of *Weak Link* where substitution is not possible and the less productive sector creates a bottleneck for production at the national level[14].

Indeed, it is preferable to use a measure that captures the underlying distribution of productivity among intermediate goods and their importance as inputs in the production of other goods. In fact, measures proposed in the previous paragraph will not capture the probability of observing these values which is given by the frequency/weight of each sector in the economy. The same minimum level of productivity can be similar between two countries, but almost never observed in one of them or its relevance as intermediate good in production might differ across countries. This will be missed by the simple use of a minimum values or simple differences of productivity which do not take into account the shape of the distribution of productivity. The construction of a measure that integrates all these aspects is covered in the following two steps.

Step 2: Kernel density of productivity and measurement of abnormal deviations

A kernel density of residual productivities ($\widehat{\mathrm{r.q}}_{c,s,t}$) is estimated for each country in each year. For all residuals productivities observed in a country in a given year, the kernel density 'spreads' the probability masses associated to the 28 (or less) sectors across $1'000$ points. All sectors in the country have an identical weight ($1/n_{c,t}$[15]) in this first estimation of the kernel density. Figure 1 shows two examples of the estimations of kernel densities for the residuals of labor productivities. The mass under the density functions is particularly large around zero, the expectancy of residual productivity. However, it is important to highlight that the distribution of residual productivities does not show a particular nor symmetric shape. The first example is for United States in 1985 and the second one is for Germany in 2000. In both graphs, the (blue) solid line is the estimated kernel density for this simple version where all sectors have the same weight. Other lines in this Figure will be commented in step 3.

[14] See Jones (2011).

[15] $n_{c,t}$ is the number of sectors observed in country c in year t.

[FIGURE 1 HERE]
[KERNEL DENSITIES OF PRODUCTIVITY ESTIMATES IN MANUFACTURES AND INTERMEDIATES]

Given that the distribution of adjusted productivities is now 'known' for each country and each period, different characteristics of the distribution can be drawn, e.g. mean and median productivity or standard deviation. The proxy for *Weak Links* proposed in this paper will capture the likelihood of observing relatively low productive sectors used as an intermediate input by other sectors in country c and year t. Therefore, a threshold for abnormal low productivities with respect to the mean productivity[16] in a country and year ($Weak_{c,t}$) is defined using λ times the standard deviation of the distribution :

$$Weak_{c,t} = \text{Prob}\left[\widehat{\text{r.q}} < mean_{c,t}(\widehat{\text{r.q}}) - \lambda \times std_{c,t}(\widehat{\text{r.q}})\right] \tag{8}$$

Results in this paper use two values for $\lambda = \{1, 1.5\}$[17] which were arbitrarily chosen as there is no clear way to determine a proper cutoff besides the obvious fact that the probability is monotonically decreasing in λ. In fact, it has to be noted that this is the first attempt to measure *Weak Links* to our knowledge and we lack of a background of potential empirical measures for this phenomenon. Therefore, the proposition above is then a primer in this sense.

However, the proxy for *Weak Links* based on kernel densities that equally weight sectors within a country does not take into account the importance of each sector as an input in the production of other sectors or the degree of substitutability/tradeability for the goods produced by each sector. In fact, each sector is given the same weight regardless of its economic relevance and the next step of our procedure aims at correcting this undesired behavior.

Step 3: Correcting sectoral importance in intermediates and tradeability/substitutability

In order to consider the relative importance of sectors as intermediate inputs, we use the share in domestic intermediate sales as weights in the kernel density estimation for each sector. However, this simple weighting would not consider the possibility of substituting domestic production by imports and therefore it is necessary to readjust the simple weights according to the level of tradeability of each sector. Specifically, we use the following (multiplicative[18]) double weights :

[16] The choice of mean productivity as the reference is due to the fact that terms B_ρ and B_θ are upper-bounded by the mean productivity. Appendix A.1 shows that these two productivity composites are particularly sensitive to observations significantly below this average value and this is the reason why we seek at capturing abnormal deviations on the left side of the distribution of productivities.

[17] Hereafter, we call moderate deviations those being one standard deviation below the mean of residual productivities and severe deviations those being 1.5 times below the mean.

[18] We decide to use a multiplicative formulation of weights that reduces the share in intermediate

$$\omega_{s,c} = \frac{\text{int sales}_{dom,s,c}}{\sum_{s=1} \text{int sales}_{dom,s,c}} \times \left[\frac{\text{int sales}_{dom,s,c}}{\text{int sales}_{tot,s,c}} \right]. \tag{9}$$

OECD input-output tables[19] are used to capture the needs of sectoral production as intermediate inputs and this information is available for 48 countries[20]. This source is privileged given similarities in the sectoral disaggregation to INDSTAT data but also because of the distinction between foreign and domestic inputs used as intermediate goods which allows to control for the substitutability by imports.

The first part of $\omega_{s,c}$ rescales the occurrence of productivity according to the production needs of the whole economy for each sector which Acemoglu et al. (2012) show to be significant determinants of fluctuations of the aggregate output of the country. The second component lowers the weights to extent that substitution of national production by imports is taking place and thus, it is a decreasing function of the level of tradeability for sector s in country c[21]. In fact, trade is particularly relevant for *Weak Links* which are a bottleneck for aggregate productivity only if substitutability of inputs is not possible. Trade might help circumventing such low productivities by replacing domestic inputs by imported ones. Thus, trade policy plays a significant role in preventing productivity losses and in every country it exists a large heterogeneity of 'protection' at the sectoral due to a variety of trade policies[22] i.e quotas, permissions, licensing, etc.

In Figure 1, the (red) dashed line corresponds to the densities estimated using sectoral shares in intermediates sales as weights for the kernel density estimations and the (green) long-dashed line adds the tradeability index to weights in the previous estimations. For the United States, shares in intermediate sales reduce the role of (very) low productive sectors by shifting the distribution to the right but the index of tradeability gives more importance to intermediate goods in the lower part of the distribution of residual productivity[23]. For Germany, the distribution of productivity is not significantly affected while using intermediate sales' shares but the index of tradeability show that domestic producers with low productivity tend to be less protected and can be more easily replaced.

Finally, the measure proposed in Equation 8 and calculated over kernel density estimations using weights in Equation 9 proxy the propensity to observe low productivity in

sales proportionally to the share of imports in the sector. Another possibility would be to consider additive weights but then, the choice of the scale while summing of weights would remain arbitrary.

[19] http://www.oecd.org/sti/inputoutput/

[20] Whenever an input-output table is not available for a country, the average structure of economies in the same region are used. The value-added share for a sector corresponds to the mean value-added share for the same sector across all countries in the region.

[21] Weights $\omega_{s,c}$ need to be normalized once the level of tradeability is used as the sum of weights is no longer equal to one.

[22] Lobbying for protection by low productive sectors would be another component of the trade policy.

[23] This might be interpreted as protectionism to low productive sectors by the United States.

important dwonstream sectors of the economy with limited possibilities for substitution through imports.

Let us move to the descriptive analysis of the proposed measure for *Weak Links*. Figure 2 shows the histograms for proxies for *Weak Links* built using the two values of λ mentioned before. Values of proxies range between 0 and 0.25 and in both cases, the most frequent interval of values is [0, 0.01[. Table 1 provides some descriptive statistics for different periods and income levels in the sample. There is no significant differences between the probabilities of observing a *Weak Link* over time. On average over the whole sample, a country has 7.6% of chances of detecting low productivity ($\lambda = 1$) in intermediates and the probability of observing extremely low productive ($\lambda = 1.5$) sectors is equal to 4.9%. 1'292 among 2'689 observations in our sample are subject to low productivity in downstream sectors i.e. probability equal or higher than 5% while controlling for most severe deviations. The bottom part of the Table shows the mean values of the proxies for four income groups of approximately the same size. The group with the highest income level shows a lower tendency to observe *Weak Links* and this is particularly lower for severe deviations of productivity. This points out that rich countries have been able to cope better with low productivities in intermediates than countries at earlier stages of development and they were able to circumvent these productivity bottlenecks to some extent. Nonetheless, extreme deviations can be observed in all income groups and there is not a clear pattern between *Weak Links* and the level of development.

[FIGURE 2 HERE]
[DISTRIBUTION OF WEAK LINKS' PROXIES]

[TABLE 1 HERE]
[SUMMARY STATISTICS OF WEAK LINKS' PROXIES]

Figures 3 and 4 show the geographical incidence of *Weak Links* and the observed heterogeneity of values within regions points out that *Weak Links* are not specific to one particular region. Furthermore, the probability of having relatively low productivity in intermediates is not restrained to developing countries and several advanced economies in Europe show high values for the two versions of the proxy[24].

[FIGURES 3 AND 4 HERE]
[AVERAGE VALUES OF WEAK LINKS' PROXIES ACROSS COUNTRIES]

[24]Figures A.5 and A.6 displays the evolution of the mean average probability through time as well as the increased coverage for recent years. The probability of observing *Weak Links* for each country varies over decadal periods and it points out that relatively low productivity is a recurrent problem through the development process of countries.

The proxy for *Weak Links* aims at capturing the effects of B_ρ in the theoretical model. However, given that domestic levels of productivities would also have an impact on B_θ through final consumption, the measure considered here captures by construction the effect of low productivity on the interaction term of B_ρ and B_θ in Equation 3. Usual concerns on the way *Weak Links* is measured in Equation 8 refer to sudden economic boom of a particular sector that might affect the shape of the distribution or the sensitivity of characteristics used in the proxy to outliers in the distribution of adjusted productivity. Manufacture data are less subject to volatility in labor productivity since volatility is mostly related to sudden and important price changes which are more likely to be observed in commodity markets. Moreover, we check for correlation between the likelihood of low productivity and the standard deviation which would be jointly evolving if an outlier would influence the shape of the distribution. Our result shows a significant but still low correlation (-0.23) between the proxy for *Weak Links* and standard deviations in the sample.

Table 2 looks further on what is being captured by the proxy. First, we want to know how frequently a sector shows a productivity below the threshold established by the measure of moderate deviations ($\lambda = 1$). Given that not all sectors are active in all countries, we have to consider the number of times a sector is detected as a low productive sector with respect to the number of times the sector is active in all country-year observations. The third column of Table 2 shows this frequency by sector[25].

Being detected as a *Weak Link* seems to be mainly related to two different types of industries; the first being related to very basic manufactures as food, beverage and tobacco and the second to the manufacture of chemicals, petroleum, coal and related products[26]. This result is due to the conditions captured by the proxy and despite the fact of lowering the importance of sectors with low intermediate sales and high tradeability, low productive sectors will always remain observable as long as sectoral weights remain positive. This is an important shortcoming of the measure that by definition captures the existence of low productive sectors regardless how intensively these sectors interact with the rest of the economy. Nonetheless, this column already provides important information concerning our proxy as it discards that the existence of a sector in an economy necessarily implies the appearance of *Weak Links*.

Remember that the proxy proposed in Equation 8 also controlled for tradeability and intermediate sales' share of sectors since it is necessary but not sufficient have low productivity to be a *Weak Link*. Columns 5 and 6 of Table 2 report the mean and median share of domestic production in total intermediate sales by sector which is the second component of the double weights used in Equation 9. Sectors where low productivity was detected in previous columns seem to be dependent to domestic production by more than three quarters of production. Only for chemicals and non-ferrous metal

[25] All descriptive statistics are presented for the proxy built using one standard deviation with respect to the mean productivity, $\lambda = 1$.

[26] Non-ferrous metal basic industries also show a high incidence of low productivity.

basic industries, imports represent 40-50 % of the needs in intermediate sales and can help circumventing low productivity at the national level.

Last column of this Table reports the average share in total intermediate sales per sector which is the mean value of the first component of weights in Equation 8. Among those sectors where relative low productivity is frequently observed, petroleum refineries concentrates a significant share of intermediate sales and it is the only sector covering the three conditions for being a *Weak Link* : (i) relative low productivity, (ii) low tradeability and (iii) intensively used as intermediate input by other sectors. It is important to notice that high values in the double weights used in the kernel density estimation do not necessarily lead to higher propensity of *Weak Links* and relatively low productivity remains the necessary condition.

[TABLE 2 HERE]
[WHAT IS BEHIND THE WEAK LINKS' PROXY ?]

How will the proxy proposed react to changes in the productivity distribution ? Consider a distribution of productivities in a given country and for simplicity suppose it resembles to a bell-shaped distribution. The fact that the productivity of a specific sector lowers in the country would have a negative impact on terms B_ρ and B_θ and therefore, their product and aggregate income would be lower. The proxy for *Weak Links* will tend to detect a higher probability in the lower tail as long as the characteristics of the distribution (mean and standard deviation) are not significantly affected by the sectoral productivity change. However, given that sectors rarely concentrate a weight larger than 5 %, the likelihood of not capturing such changes is low. The increase in probability measured by the proxy implies a lower level of aggregate income and thus, the expected estimate for the proxy included in Equation 3 instead of B_θ and B_ρ is negative. Next section explores the growth effects of the probability of observing low productive sectors for a large sample of countries.

4 Weak Links and Growth: any costs ?

It would have been suitable to confront the presence of *Weak Links* with common growth determinants[27] in the literature in order to create a comparative analysis, but scarcity of complete series for those determinants is a major impediment. In fact, two thirds of the sample for which the *Weak Links* proxy is calculated would not be considered if other determinants are included in the regression analysis and in terms of policy perspectives, it would not be straightforward to extend conclusions to developing countries given that the empirical evidence would mainly rely on countries with higher data quality.

[27] See Barro (2000).

In this empirical analysis of the impact of *Weak Links* on aggregate growth, we use GDP per capita in constant prices from the World Development Indicators for the level of development of countries and Table 3 presents the pairwise correlations of *Weak Links* with common growth determinants and the level of development of countries. All correlations have the expected sign. The correlation of the two *Weak Link* proxies considered here and the level of development is negative. Indeed, the existence of a *Weak Link* necessarily implies a loss of productivity which is spread all over the economy and sums up at the aggregate level. Moreover, losses of efficiency by reallocation of resources in order to counter the *Weak Link* might generate further distortions as a suboptimal use of resources which reinforces the negative impact on the level of development.

[TABLE 3 HERE]
[CROSS CORRELATION TABLE]

Given that the usual regression in income levels cannot be implemented here, we use a first-differences approach as done by Wacziarg and Welch (2008) while testing the impact of liberalization on growth and the most complete version of our specification is the following :

$$\Delta \ln GDP_{c,t} = \Delta \, Weak\ Links_{c,t} + \sum_c \delta_c + \sum_t \gamma_t + \nu_{c,t} \tag{10}$$

where $\Delta GDP_{c,t}$ stands for the annual growth rate of GDP per capita of country c in year t in thousands US$ at constant prices, $\Delta\, Weak\ Link$ are changes in the probability of observing relative low productivities and δ_c and ι_t are country and time fixed effects.

The first differences approach allows us to neutralize most of country characteristics that remain relatively stable over time and to keep the complete sample for which the proxy for low productivity in intermediate goods has been calculated. Table A.1 provides the number of years by country, over the period 1963-2001, for which we calculated a value for the proxy for *Weak Links*. Given that our analysis is run on first differences, countries with only one observation are dropped by default and the analysis is based on 119 countries with 22 observations on average. One third of countries in the sample has at least 30 observations and another third has at least 16 observations. Among the 40 countries with the largest number of observations, half of them are developing countries such as Ecuador, India, Malaysia or Zimbabwe. Few countries have limited number of observations but this is also due to reunification processes such as Yemen or Germany and the formation of new countries like Croatia or the Czech Republic.

Table 4 shows the results obtained by regressing annual growth rates on changes on the likelihood of *Weak Links*. Annual growth rates are used aiming to keep a coherent and uniform endogenous variable instead of differences of income levels (in log) because series are not continuous over time for all countries. Aiming at learning about the magnitude of the impact of low productivities, we first introduce the proxy measuring the probability of observing moderate deviations (one standard deviation below the mean, $\lambda = 1$) and then, the proxy for severe deviations (one and a half standard deviations to the mean, $\lambda = 1.5$). Increases in the probability of relatively low productive sectors have a negative impact of annual growth rates and this result is robust to the inclusion of year fixed effects which combined with the country fixed effects control for a large set of characteristics such as factor endowments of each economy and global external shocks. The estimated impact is more pronounced when the measure controlling for larger deviations is used. Undoubtedly, severe deviations would imply higher costs due to complementary in the use of intermediate inputs.

[TABLE 4 HERE]
[GROWTH REGRESSIONS]

Given the novelty of the analysis conducted here, we lack of guidance on how the proxy for *Weak Links* should be considered in Equation 10. Thus, a second version of the specification is tested to check the robustness of results. The explanatory variable is no longer the change in the probability but the change in logs of the probability of observing a *Weak Link* plus one ($\ln P_{-\sigma} = \ln(\mathit{Weak}(-\sigma)+1)$). This second specification allow us to interpret estimates for the *Weak Links'* proxy as elasticities instead of semi-elasticities as in the first specification. Results are consistent across these two specifications of the regression and show similar magnitudes of estimates with respect to the extent of deviations.

Tables A.2 and A.3 in the Appendix address potential concerns on our explained variable which use observations not uniformly spanned over time while creating annual growth rate of per capita income. The first control consists of a dummy variable taking the value 1 whenever the growth is not calculated on two consecutive observations over time and zero otherwise. The second one includes the span of time between the two observations as a control. In almost all cases, estimates for these control variables do not reach statistically significance. We conclude that results discussed previously are not driven by any lacunary collection of information.

Table 5 splits our sample for selected thresholds on the level of development. The first split is around 3 thousand US$ of income per capita which corresponds to the median value of the variable. Severe deviations of productivity in richest countries is the only specification where the estimate of the proxy achieve statistical significance. While splitting at 10 thousand US$, three of four coefficients on *Weak Links'* proxies are significant. Indeed, the lack of significance among the poorest countries (below 3K US$) is certainly related to the complexity of the production in these countries. If most of national production is generated by a single export-oriented sector as agriculture or mining, the dissemination of low productivity in intermediate goods is almost unnoticed given that the country does not create a complex chain of production but it is devoted to the extraction or production of a unique good. In such countries, growth performance is determined by the evolution of the most relevant sector in the economy.

Finally, results in this Table point out two important facts: moderate productivity deviations in intermediate goods have a significant impact in economies having reached a certain level of development (above 10K US$) and more importantly, severe deviations matter for all developing countries. These latter deviations impede development by reducing chances to develop industries with higher value-added that might lift people out of poverty. Addressing the need for productivity improvements is another tool to foster growth and should be carefully considered by policymakers in a coherent framework of industrialization.

[TABLE 5 HERE]
[PRODUCTIVITY LOSSES IN DEVELOPING AND DEVELOPED COUNTRIES]

4.1 Beyond short-term results

All previous results show the short-term effects of an increase of the likelihood of observing low productivity and this section analyzes its impact on medium-term growth. Table 6 is the first try to evaluate the impact on annual growth rates calculated over longer periods of time, i.e three and five years. Given the discontinuity of series for several countries, picking selected years would reduce significantly the sample size. Instead of that, the following procedure to span observations over time for each country is implemented. Once ordered over time, the first observation of each country (t_0^c) is kept and given the time span imposed between observations (τ), all observations in each country that belong to the interval $]t_0^c, t_0^c + \tau[$ are dropped. Separately by country, the first of remaining observations is denoted t_1^c and the procedure is repeated until the last observation of each country.

[TABLE 6 HERE]
[GROWTH REGRESSIONS FOR LONGER PERIODS]

Regardless the time span selected country-observations, severe drops of sectoral productivity have an impact on growth whereas moderate drops do not have a statistically significant effect. Its estimate ranges between -0.088 and -0.074 and in the case of a decrease of 10% in the probability of observing a low productivity in intermediate goods, growth of income per capita would be boosted by 0.83% annually. Suppose an economy where the 5-year growth has been near to zero, the decrease mentioned before would have ensured a 4.2% ($= 1.0083^5 - 1$) growth of the income per capita over the period.

Table A.4 shows the results while picking selected year-observations for the empirical analysis. This procedure which is not data driven as the one presented before generates a different sample for testing the impact of *Weak Links* on medium-term growth. Results are consistent between those in Table 6 and once again, most marked deviations of productivities have the most of impact on growth.

4.2 Endogeneity of trade policy/protection

Low productive sectors might try to obtain protective measures from the government such as trade restrictions which allow them to counter substitution by imports and ensure their survival despite extremely high costs of production[28]. If protectionism is an outcome of lobbying activities at the sectoral level, endogeneity of protection through trade policy might explain to some extent previous results. Trade policy will be driven by sectoral productivity levels and trade policy is relevant in explaining the growth pattern followed by countries. Most of papers find that outward-oriented countries have a better growth perfomance (Dollar (1992), Frankel et al. (1999)) and more recently, Wacziarg et Welch (2008) show that liberalization has a positive impact on growth. Consequently, our measure would partially capture effects related to trade restrictiveness and not only pure effects of sectoral productivity and complementarity.

Table 7 addresses this issue using the instrumental variable approach where our explanatory variable is instrumented by an analogous proxy which ignores levels of tradeability in intermediate sales while defining weights $\tilde{\omega}_{c,s}$ for the kernel density of residual productivities and which is not affected by the potential endogeneity of trade policy. The instrumental variable is called $Weak_{notrade}$. A second instrumental variable ($Weak_{noweight}$) is the proxy obtained without using weights in the estimation of the kernel density. This is equivalent to consider that intermediate sales of all sectors are equally important.The first stage in all specifications is the following linear regression estimated by OLS:

$$\begin{aligned} \Delta\, Weak\ Links_{c,t} \;=\; & \beta_1 \Delta\, Weak\ Links_{notrade,c,t} + \beta_2 \Delta\, Weak\ Links_{noweight,c,t} \\ & + \textstyle\sum_c \mu_c + \sum_t \phi_t + \xi_{c,t}. \end{aligned} \tag{11}$$

The two stages of the IV procedure for all the specifications are shown in Table 7. The instrumental variable that ignores the tradeability of intermediate goods is the most significant instrument in predicting our proxy for *Weak Links* and the reason for it relies on the fact that the consumption structure of intermediate inputs is a relevant information while trying to estimate the effects of complementary in production. This result is in line with the findings by Acemoglu et al. (2012) which point out that

[28] Jones (2011) considers extractive distortions at the sectoral level but his framework is less adapted to consider 'preferences' granted to sectors such as protection from import competition.

linkages across sectors determine the transmission of shocks from the sectoral level to the aggregate level. Estimates in the second stage are statistically significant and of similar magnitudes to our previous results. The estimates here confirm the effects found previously in Table 4[29].

[TABLE 7 HERE]
[ADDRESSING TRADE POLICY ENDOGENEITY]

Results concerning medium-term effects of *Weak Links* on growth are also confirmed using the 3-year and 5-year span between observations. Instrumental variables are the same as those used in Table 7 and results are presented in Tables 8 and 9. Nonetheless, the exogeneity of instruments is often rejected for the proxy controlling for severe deviations when year fixed effects are not included in the estimation procedure.

[TABLES 8 AND 9 HERE]
[ADDRESSING TRADE POLICY ENDOGENEITY (GROWTH SPELLS OVER 3 AND 5 YEARS]

5 Conclusions

This paper aimed at integrating the rationale developed by Jones (2011) in an empirical exercise enabling to quantify the impact of *Weak Links* on growth. In that sense, we proposed a measure that controls for low productivity in intermediate goods by taking into account the production structure of economies and the heterogeneous level of tradeability at the sectoral level. The measure presented here uses information based on the underlying distribution of productivity by country and year and we use a rich dataset for 119 countries around the world to estimate a probability of observing *Weak Links*. Furthermore, we are able to identify crucial sectors that might generate important productivity losses at the aggregate level by analyzing the characteristics of the proxy.

In a second stage, we use two versions of the proxy for moderate and severe deviations in productivity to assess their impact on aggregate growth. We found that changes on the probability of observing *Weak Links* have a significant impact on annual growth rates. As expected, severe deviations are those having a larger impact on development perspectives and even though developing countries seem to be less affected due to their simplified economic structure, *Weak Links* do matter for the development of all countries.

[29]The test for weak instruments and for the exogeneity of instruments validate our results.

For instance, severe deviations in terms of relative productivity do not only affect short-term growth rates but have a significant impact on growth perspectives for longer periods (3- and 5-year periods). A decrease of 10 percentage points in the probability of observing severe *Weak Links* implies a boost of 1% in annual growth. These results are consistent across different specifications and we use instrumental variables to control for endogeneity concerns related to trade policy.

The empirical evidence in this paper points out the need of an inclusive and coherent development strategy by policymakers. Such strategy should prevent abnormal productivity deviations particularly for goods intensively used as intermediate inputs. Promoting productivity improvements would generate large and positive spillovers across the economy. Finally, trade policy but also public investment are relevant policies to address the potential shortcoming of a productivity bottleneck and policymakers should carefully consider their use.

Tables and Figures

Fig. 1: Kernel densities of productivity estimates in manufactures and intermediates

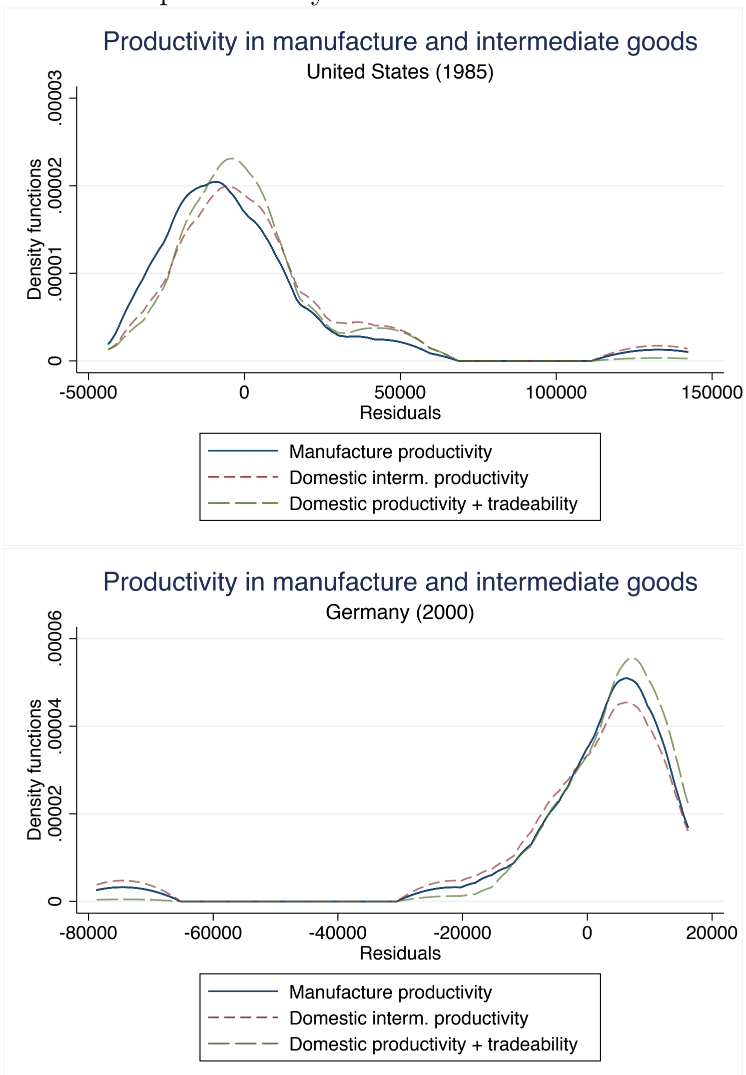
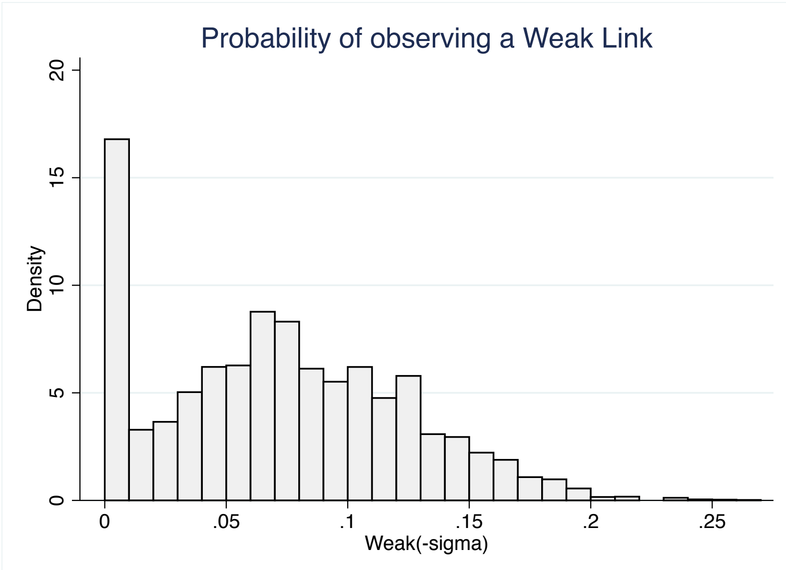


Fig. 2: Distribution of the *Weak Links'* proxies



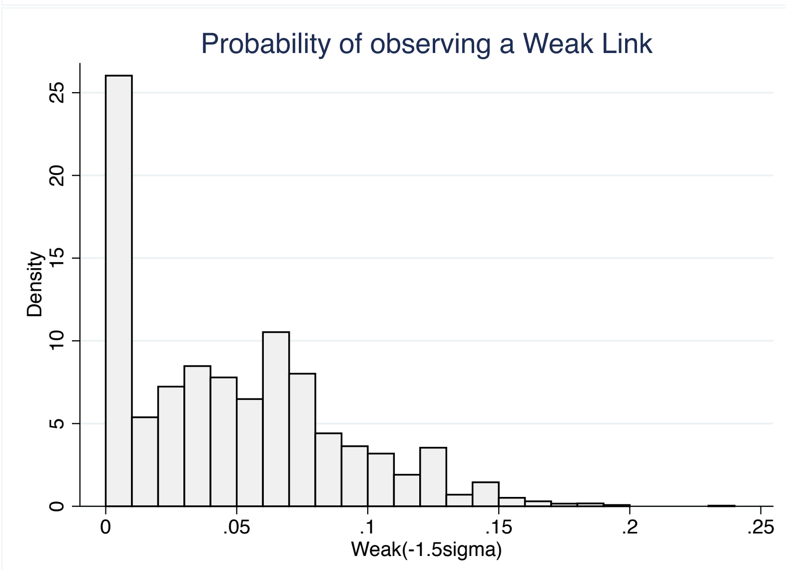
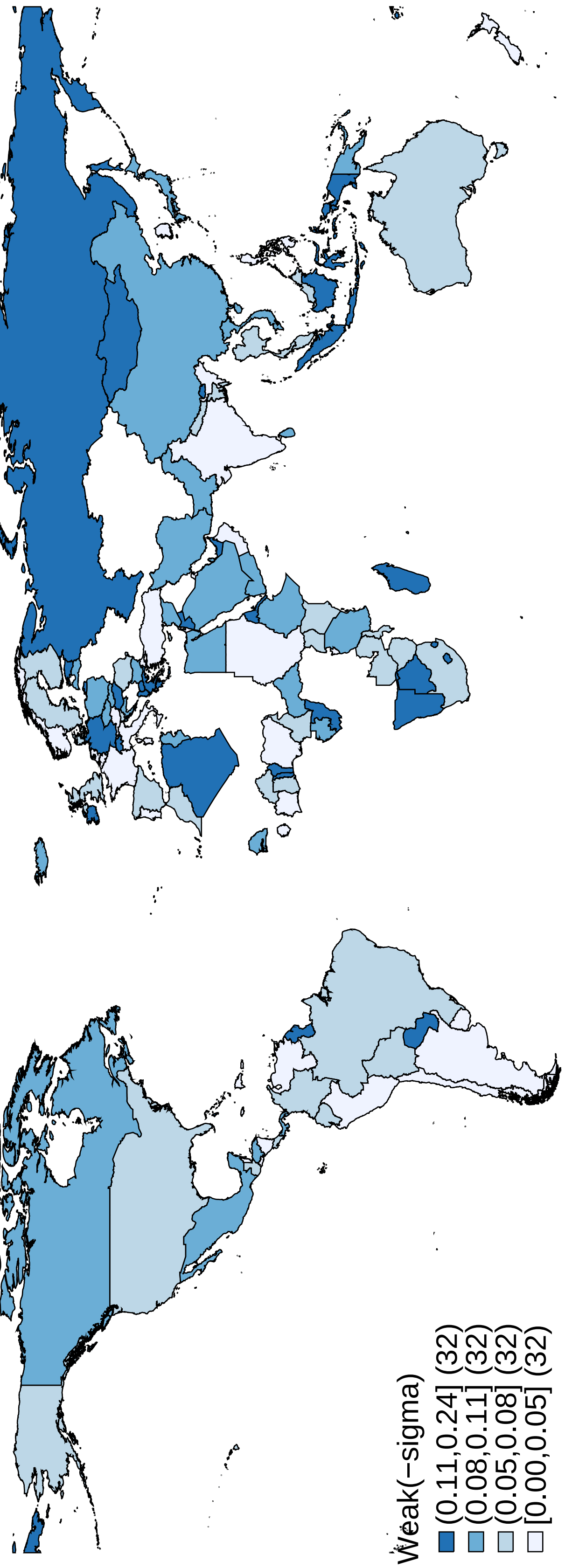
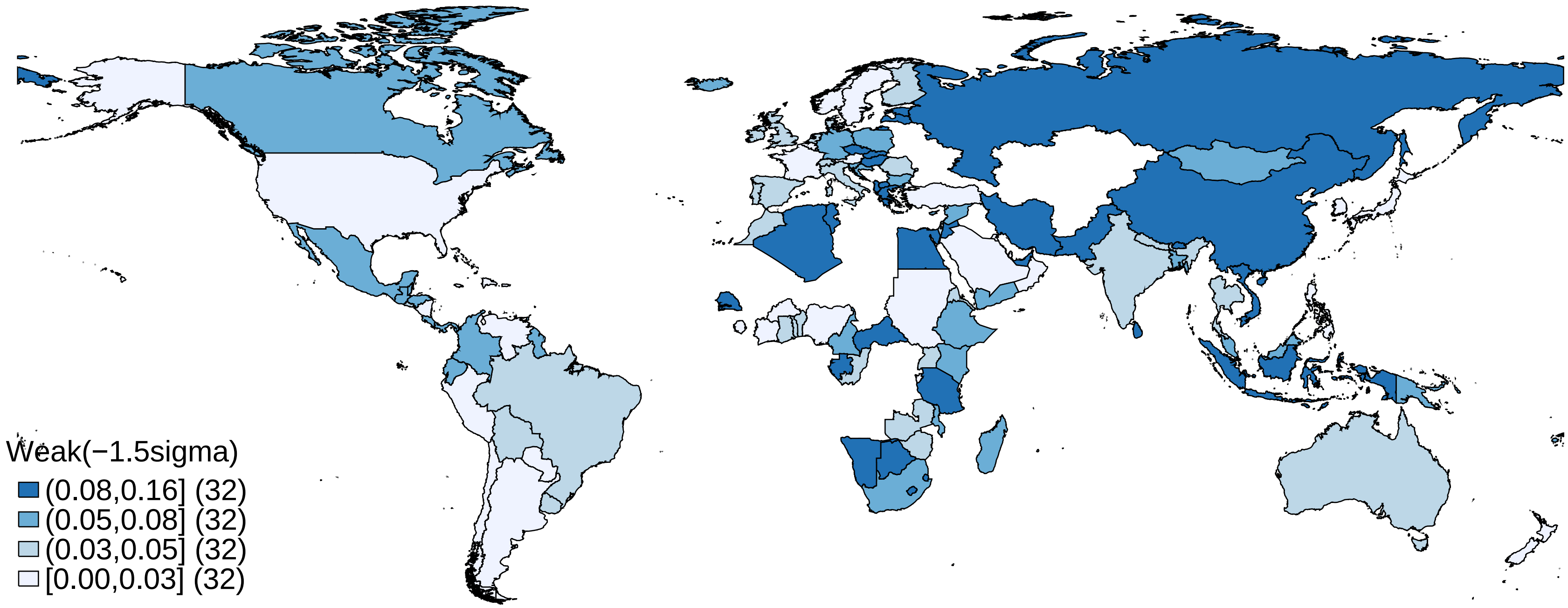


Fig. 3: Average value of *Weak Links'* proxy across countries



Note: The *Weak Links'* proxy used in this Figure is the probability of observing productivities lower than the mean productivity minus the standard deviation averaged over all observations for each country in the period 1963-2001.

Fig. 4: Average value of *Weak Links'* proxy across countries (bis)



Note: The *Weak Links'* proxy used in this Figure is the probability of observing productivities lower than the mean productivity minus 1.5 times the standard deviation averaged over all observations for each country in the period 1963-2001.

Tab. 1: Summary statistics of *Weak Links'* proxies

	Mean	Std. Dev.	Min.	Max.
	in 1960s			
Observations	372			
Weak Links(-1.5σ)	0.051	0.038	0	0.189
Weak Links(-σ)	0.079	0.048	0	0.190
	in 1970s			
Observations	736			
Weak Links(-1.5σ)	0.048	0.040	0	0.160
Weak Links(-σ)	0.070	0.049	0	0.211
	in 1980s			
Observations	812			
Weak Links(-1.5σ)	0.048	0.041	0	0.173
Weak Links(-σ)	0.077	0.054	0	0.253
	in 1990s			
Observations	898			
Weak Links(-1.5σ)	0.049	0.044	0	0.236
Weak Links(-σ)	0.078	0.055	0	0.260
	$GDP_{pc} < 1'000$ USD			
Observations	741			
Weak Links(-1.5σ)	0.057	0.040	0	0.195
Weak Links(-σ)	0.082	0.047	0	0.260
	$1'000$ USD $< GDP_{pc} < 3'000$ USD			
Observations	608			
Weak Links(-1.5σ)	0.059	0.049	0	0.178
Weak Links(-σ)	0.080	0.056	0	0.236
	$3'000$ USD $< GDP_{pc} < 12'000$ USD			
Observations	702			
Weak Links(-1.5σ)	0.048	0.039	0	0.189
Weak Links(-σ)	0.071	0.053	0	0.253
	$GDP_{pc} > 12'000$ USD			
Observations	638			
Weak Links(-1.5σ)	0.029	0.028	0	0.131
Weak Links(-σ)	0.065	0.050	0	0.202

Observations from 2001 and 2000 are included among observations in 1990s.

Tab. 2: What is behind the *Weak Links'* proxy ?

ISIC Sector	Description of the sector	Low prody Freq.	Low prody Obs.	Tradeability Median	Tradeability Mean	Share Int. Sales
311	Food manufacturing	1.3	35	**82.0**	**81.1**	**6.6**
313	Beverage industries	**15.4**	**414**	**82.0**	**81.1**	2.5
314	Tobacco manufactures	**53.4**	**1'286**	**82.0**	**81.1**	1.7
321	Manufacture of textiles	2.3	64	62.7	60.6	1.4
322	Manufacture of wearing apparel, except footwear	6.3	162	62.7	60.6	0.7
323	Manufacture of leather, except footwear and wearing apparel	3.9	95	62.7	60.6	0.2
324	Manufacture of footwear	5.4	129	62.7	60.6	0.2
331	Manufacture of wood	2.2	60	**81.8**	**78.4**	4.6
332	Manufacture of furniture and fixtures	1.7	44	**76.0**	**74.6**	1.6
341	Manufacture of paper	1.0	27	**80.3**	**76.5**	**5.5**
342	Printing, publishing and allied industries	0.4	11	**80.3**	**76.5**	**7.0**
351	Manufacture of industrial chemicals	**9.8**	**262**	53.9	53.4	5.1
352	Manufacture of other chemical products	6.1	153	53.9	53.4	**6.3**
353	Petroleum refineries	**63.8**	**1'285**	**76.4**	**69.4**	**7.3**
354	Manufacture of miscellaneous products of petroleum and coal	**29.9**	**424**	**76.4**	**69.4**	1.9
355	Manufacture of rubber products	4.0	100	68.8	66.1	3.0
356	Manufacture of other plastic products	0.5	12	68.8	66.1	3.9
361	Manufacture of pottery, china and earthenware	8.8	199	**85.1**	**82.1**	1.5
362	Manufacture of glass and glass products	1.4	31	**85.1**	**82.1**	1.5
369	Manufacture of other non-metallic mineral products	0.6	16	**85.1**	**82.1**	**6.7**
371	Iron and steel basic industries	4.1	95	61.9	60.9	**7.4**
372	Non-ferrous metal basic industries	**12.2**	**234**	61.9	60.9	3.1
381	Manufacture of fabricated metal products	0.6	17	**80.8**	**76.0**	**9.7**
382	Manufacture of machinery except electrical	0.9	22	56.7	53.6	5.4
383	Manufacture of electrical machinery	0.7	18	41.7	45.1	5.3
384	Manufacture of transport equipment	3.1	80	53.5	50.0	3.4
385	Manufacture of professional, scientific, measuring and controlling equip.	4.4	90	38.8	41.2	1.2
390	Other Manufacturing Industries	2.2	55	**76.0**	**74.6**	1.2

Low productivity counts how many times a sector is detected as a low productive sector and the frequency of those events given all observations in the sample for that sector. Tradeability measures the share of national production in intermediate sales by sector across countries and it corresponds to the second part of the weight used in the estimation of our proxy. Share in intermediate sales is the average share in total intermediate sales for goods produced by the sector and it corresponds to the first part of ω in the kernel density estimation.

Tab. 3: Cross-correlation table

	GDP	$\ln GDP$	Second. enroll.	Prim. enroll.	Gover. consump.	Pop. density	*Weak* $(-\sigma)$	*Weak* (-1.5σ)
GDP	1.00							
$\ln GDP$	0.85	1.00						
Secondary enrollment	0.68	0.82	1.00					
Primary enrollment	0.17	0.41	0.45	1.00				
Government consumption in GDP	0.26	0.25	0.29	-0.05	1.00			
Population density	0.15	0.17	0.04	0.03	-0.17	1.00		
Weak$(-\sigma)$	-0.11	-0.13	-0.07	-0.09	0.05	0.05	1.00	
Weak(-1.5σ)	-0.29	-0.26	-0.19	-0.06	-0.00	-0.06	0.72	1.00

Tab. 4: Growth regressions

$\Delta \ln GDP$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ *Weak* $(-\sigma)$	-0.046*				-0.040*			
	(0.025)				(0.024)			
Δ *Weak* (-1.5σ)		-0.085**				-0.083**		
		(0.035)				(0.034)		
$\Delta \ln P_{-\sigma}$			-0.053**				-0.046*	
			(0.027)				(0.026)	
$\Delta \ln P_{-1.5\sigma}$				-0.091**				-0.089**
				(0.037)				(0.036)
Constant	0.021***	0.021***	0.021***	0.021***	0.047***	0.048***	0.047***	0.048***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)	(0.007)	(0.007)	(0.007)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE					✓	✓	✓	✓
Observations	2689	2689	2689	2689	2689	2689	2689	2689
R^2	0.171	0.172	0.172	0.172	0.240	0.241	0.240	0.241

$\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. Δ *Weak* $(-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tab. 5: Productivity losses in developing and developed countries

$\Delta \ln GDP$	Splitting around 3K US$ Below		Splitting around 3K US$ Above		Splitting around 10K US$ Below		Splitting around 10K US$ Above	
Δ *Weak* $(-\sigma)$	-0.035		-0.041		-0.022		-0.053*	
	(0.039)		(0.027)		(0.031)		(0.032)	
Δ *Weak* (-1.5σ)		-0.083		-0.099**		-0.072*		-0.116**
		(0.051)		(0.044)		(0.041)		(0.056)
Constant	0.018	0.018	0.038***	0.039***	0.050***	0.051***	0.050***	0.050***
	(0.012)	(0.012)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.012)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1349	1349	1340	1340	1932	1932	757	757
R^2	0.215	0.216	0.338	0.340	0.242	0.243	0.361	0.362

$\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. Δ *Weak* $(-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tab. 6: Growth regressions for longer periods

Annual growth over 3-year periods						
$\Delta \ln GDP$	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
Δ *Weak* $(-\sigma)$	-0.022		-0.006		-0.005	
	(0.024)		(0.023)		(0.023)	
Δ *Weak* (-1.5σ)		-0.096***		-0.079***		-0.083***
		(0.032)		(0.030)		(0.030)
Constant	0.021***	0.020***	0.051**	0.050**	0.018***	0.018***
	(0.001)	(0.001)	(0.023)	(0.023)	(0.004)	(0.004)
Country FE	✓	✓	✓	✓	✓	✓
Year FE			✓	✓		
Period FE					✓	✓
Observations	879	879	879	879	879	879
R^2	0.308	0.316	0.422	0.427	0.389	0.395

Annual growth over 5-year periods						
$\Delta \ln GDP$	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
Δ *Weak* $(-\sigma)$	-0.032		-0.022		-0.020	
	(0.024)		(0.023)		(0.023)	
Δ *Weak* (-1.5σ)		-0.083**		-0.088***		-0.074**
		(0.033)		(0.031)		(0.031)
Constant	0.021***	0.021***	0.015	0.012	0.022***	0.021***
	(0.001)	(0.001)	(0.022)	(0.022)	(0.003)	(0.003)
Country FE	✓	✓	✓	✓	✓	✓
Year FE			✓	✓		
Period FE					✓	✓
Observations	518	518	518	518	518	518
R^2	0.421	0.428	0.577	0.585	0.521	0.527

$\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. Δ *Weak* $(-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tab. 7: Addressing trade policy endogeneity

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Stage of IV procedure	2nd	1st	2nd	1st	2nd	1st	2nd	1st
Explained variable	$\Delta \ln GDP$	Δ *Weak* $(-\sigma)$	$\Delta \ln GDP$	Δ *Weak* (-1.5σ)	$\Delta \ln GDP$	Δ *Weak* $(-\sigma)$	$\Delta \ln GDP$	Δ *Weak* (-1.5σ)
Δ *Weak* $(-\sigma)$	-0.047*				-0.041*			
	(0.024)				(0.024)			
Δ *Weak* (-1.5σ)			-0.078**				-0.076**	
			(0.036)				(0.035)	
Δ *Weak* $(-\sigma)_{notrade}$		0.962***				0.963***		
		(0.005)				(0.005)		
Δ *Weak* $(-\sigma)_{noweight}$		0.027***				0.027***		
		(0.006)				(0.006)		
Δ *Weak* $(-1.5\sigma)_{notrade}$				0.997***				0.997***
				(0.005)				(0.005)
Δ *Weak* $(-1.5\sigma)_{noweight}$				0.004				0.003
				(0.006)				(0.006)
Constant	-0.023	-0.000	0.086*	-0.000	-0.022	0.002	-0.017	0.002**
	(0.044)	(0.000)	(0.045)	(0.000)	(0.043)	(0.001)	(0.043)	(0.001)
F-statistics		3.8e+04		2.6e+04		3.7e+04		2.6e+04
Sargan test (p-value)		0.251		0.428		0.432		0.319
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE					✓	✓	✓	✓
Observations	2689	2689	2689	2689	2689	2689	2689	2689
R^2	0.171	0.968	0.172	0.954	0.240	0.968	0.241	0.955

Columns (Xa) correspond to the second stage of the IV procedure and columns (Xb) to the first stage. $\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. Δ *Weak* $(-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tab. 8: Addressing trade policy endogeneity (growth spells over 3 years)

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
Stage of IV procedure	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st
Explained variable	$\Delta \ln GDP$	$\Delta Weak\ (-\sigma)$	$\Delta \ln GDP$	$\Delta Weak\ (-1.5\sigma)$	$\Delta \ln GDP$	$\Delta Weak\ (-\sigma)$	$\Delta \ln GDP$	$\Delta Weak\ (-1.5\sigma)$	$\Delta \ln GDP$	$\Delta Weak\ (-\sigma)$	$\Delta \ln GDP$	$\Delta Weak\ (-1.5\sigma)$
$\Delta Weak\ (-\sigma)$	-0.016				-0.002				-0.000			
	(0.023)				(0.021)				(0.021)			
$\Delta Weak\ (-1.5\sigma)$			-0.084***				-0.066**				-0.068**	
			(0.030)				(0.028)				(0.029)	
$\Delta Weak\ (-\sigma)_{notrade}$		0.967***				0.971***				0.968***		
		(0.010)				(0.010)				(0.010)		
$\Delta Weak\ (-\sigma)_{noweight}$		0.015				0.015				0.016		
		(0.011)				(0.011)				(0.011)		
$\Delta Weak\ (-1.5\sigma)_{notrade}$				1.011***				1.010***				1.010***
				(0.009)				(0.009)				(0.009)
$\Delta Weak\ (-1.5\sigma)_{noweight}$				-0.010				-0.007				-0.006
				(0.011)				(0.011)				(0.011)
Constant	-0.089***	-0.000	0.007	-0.000	0.049	-0.006	0.047	-0.003	0.028	0.000	0.040	0.001
	(0.030)	(0.000)	(0.029)	(0.000)	(0.037)	(0.007)	(0.036)	(0.005)	(0.028)	(0.001)	(0.028)	(0.001)
F-statistics		550.5		295.1		505.9		289.1		533.9		298.6
Sargan test (p-value)		0.144		0.023		0.358		0.168		0.406		0.091
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE					✓	✓	✓	✓				
Period FE									✓	✓	✓	✓
Observations	879	879	879	879	879	879	879	879	879	879	879	879
R^2	0.308	0.969	0.316	0.965	0.422	0.971	0.427	0.966	0.389	0.970	0.395	0.965

Columns (Xa) correspond to the second stage of the IV procedure and columns (Xb) to the first stage. $\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. $\Delta Weak\ (-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tab. 9: Addressing trade policy endogeneity (growth spells over 5 years)

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
Stage of IV procedure	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st
Explained variable	$\Delta \ln GDP$	$\Delta\, Weak\ (-\sigma)$	$\Delta \ln GDP$	$\Delta\, Weak\ (-1.5\sigma)$	$\Delta \ln GDP$	$\Delta\, Weak\ (-\sigma)$	$\Delta \ln GDP$	$\Delta\, Weak\ (-1.5\sigma)$	$\Delta \ln GDP$	$\Delta\, Weak\ (-\sigma)$	$\Delta \ln GDP$	$\Delta\, Weak\ (-1.5\sigma)$
$\Delta\, Weak\ (-\sigma)$	-0.032				-0.022				-0.019			
	(0.022)				(0.020)				(0.020)			
$\Delta\, Weak\ (-1.5\sigma)$			-0.081***				-0.083***				-0.069**	
			(0.030)				(0.027)				(0.027)	
$\Delta\, Weak\ (-\sigma)_{notrade}$		0.957***				0.960***				0.957***		
		(0.012)				(0.012)				(0.012)		
$\Delta\, Weak\ (-\sigma)_{noweight}$		0.041***				0.040***				0.044***		
		(0.013)				(0.013)				(0.013)		
$\Delta\, Weak\ (-1.5\sigma)_{notrade}$				1.009***				1.024***				1.012***
				(0.014)				(0.014)				(0.014)
$\Delta\, Weak\ (-1.5\sigma)_{noweight}$				0.013				0.003				0.011
				(0.017)				(0.017)				(0.016)
Constant	0.020	-0.000	0.056**	-0.000	0.073**	0.000	0.073**	0.003	-0.020	-0.003**	0.028	-0.002**
	(0.024)	(0.000)	(0.023)	(0.000)	(0.030)	(0.008)	(0.029)	(0.007)	(0.022)	(0.001)	(0.022)	(0.001)
F-statistics		323.1		193.1		274.2		170.3		312.3		192.7
Sargan test (p-value)		0.416		0.032		0.930		0.124		0.839		0.067
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE					✓	✓	✓	✓				
Period FE									✓	✓	✓	✓
Observations	518	518	518	518	518	518	518	518	518	518	518	518
R^2	0.421	0.980	0.428	0.964	0.577	0.982	0.585	0.968	0.521	0.980	0.527	0.965

Columns (Xa) correspond to the second stage of the IV procedure and columns (Xb) to the first stage. $\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. $\Delta\, Weak\ (-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

A Appendices

A.1 Jones' model with linkages and complementarities

Jones (2011) models an economy where a continuum of goods is produced using physical capital (K), human capital (H) and intermediate goods (X). All goods produced by the economy can be used as intermediate goods for the production of other goods or as a final consumption good and the production of good i is given by the following Cobb-Douglas function:

$$Q_i = A_i(K_i^{\alpha} H_i^{1-\alpha})^{1-\sigma} X_i^{\sigma}, \tag{A.1}$$

where σ is the share of intermediate goods used in the production of good i and this share is constant across sectors. So long as $\sigma > 0$, low productivity in one sector is conveyed from one sector to the rest of the economy through intermediate goods' consumption. All production of each good i is allocated between the two purposes mentioned before :

$$Q_i = c_i + z_i. \tag{A.2}$$

Gross domestic production (Y) is the aggregation of all final uses and represents a single final good:

$$Y = \left(\int_0^1 c_i^{\theta} di \right)^{1/\theta} \tag{A.3}$$

and intermediate goods used by all sectors are aggregated as follows:

$$X = \left(\int_0^1 z_i^{\rho} di \right)^{1/\rho}. \tag{A.4}$$

$1/(1-\theta)$ is the elasticity of substitution between final goods and it is assumed to be higher than one which implies that $0 < \theta < 1$ and $1/(1-\rho)$, the elasticity of substitution between intermediate goods in production, is less than one ($\rho < 0$). The condition imposed on the latter elasticity of substitution can be justified by the fact that the possibility of substitutability between intermediate goods is limited while substitutability is higher in consumption.

There is an exogenous level of human capital per worker ($\bar{h}$) and the amount of human capital (H) is constrained by the number of workers ($\bar{L}$). Physical capital is a temporal constraint but it is possible to cumulate it over periods as usual and preferences are standard :

$$U = \int_0^\infty e^{-\lambda t} u(C_t) dt. \tag{A.5}$$

For a given level of human capital and physical capital, the expression of the domestic gross production can be solved for. The expression of the GDP using a symmetric allocation of resources where each sector i uses the same amount of human capital and physical capital and a constant proportion of their production is used as intermediate inputs[30] is given by :

$$Y = \phi(\bar{z})(S_\theta^{1-\sigma} S_\rho^\sigma)^{\frac{1}{1-\sigma}} K^\alpha H^{1-\alpha} \tag{A.6}$$

$$\text{with} \quad S_\rho = \left(\int_0^1 A_i^\rho di\right)^{\frac{1}{\rho}}, \quad S_\theta = \left(\int_0^1 A_i^\theta di\right)^{\frac{1}{\theta}} \quad \text{and} \quad \phi(\bar{z}) = \left((1-\bar{z})^{1-\sigma} \bar{z}^\sigma\right)^{\frac{1}{1-\rho}}.$$

σ measures the importance of intermediate inputs in the production function and α is the capital's share of output if linkages between sectors do not exist ($\sigma = 0$).

Here, it is important to undestand the behavior of productivity composites, S_θ and S_ρ, since both are defined using the same aggregator but with different values for the exponent. S_ρ has negative values for the exponent ranging between $-\infty$ and 0 while the exponent for S_θ is positive and ranges between 0 and ∞. These aggregator are monotonically increasing in values of the exponent and thus, S_ρ is always lower than S_θ. The previous ranges for exponents define the range for the values of these productivity composites whose most extreme cases are presented here below.

If substitution of intermediate goods is not possible ($\rho \to -\infty$), the total factor productivity of the country depends crucially on the lowest productivity across all sectors of the economy through S_ρ which corresponds to a sort of Leontief-type aggregate production function. S_θ is comprised between the geometric and the arithmetic means of sectoral productivities and when substituability is the lowest in consumption and intermediates, the TFP of the country depends on the product of the geometric mean and the minimum productivity. When substituability is the highest, the TFP of the country is the product of the geometric and the arithmetic mean of productivities.

However, the illustration for the case of low substituability seems to be an extreme example as competitive allocation of resources between sectors would lead to more efficient utilization of resources than relying on the minimum productivity level. Jones (2011) considers a competitive allocation of resources which leads to higher levels of GDP as productivity composites are less subject to low values of productivity.

His model also integrates distortions at the micro-level which might be related to taxes, theft or market regulations and each sector loses a fraction τ_i of its production which

[30] All sectors use the same composite of intermediate goods, the share of production of each good that is used as intermediate good is equal to $\bar{z}$ and savings are fixed.

implies a loss in productivity given sectoral expropriation rates. These micro-level distortions in the economy affect complementarities at the aggregate level and reduce the TFP of the country. However, we consider $\forall \tau_i = 0$ for all sectors as our analysis aims at capturing the impact of productivity spillovers across sectors and we do not have access to detailed data that allow us to distinguish extractive distortions at the sectoral level across countries and time periods. :

$$Y = \psi(\tau)(B_{\theta}^{1-\sigma} B_{\rho}^{\sigma})^{\frac{1}{1-\sigma}} K^{\alpha} H^{1-\alpha} \tag{A.7}$$

where

$$B_{\rho} = \left(\int_0^1 A_{)}^{\frac{\rho}{1-\rho}} di \right)^{\frac{1-\rho}{\rho}}, \tag{A.8}$$

$$B_{\theta} = \left(\int_0^1 A_i^{\frac{\theta}{1-\theta}} di \right)^{\frac{1-\theta}{\theta}}, \tag{A.9}$$

$$\text{and} \quad \psi(\sigma) = (1-\sigma)\sigma^{\frac{\sigma}{1-\sigma}}. \tag{A.10}$$

So, the unique difference between expressions A.6 and A.7 is given by the replacement of S_{θ} by B_{θ} and S_{ρ} by B_{ρ}. The formulation under A.7 imposes a further condition on θ in order to ensure that the elasticity of substitution is positive across final goods $(0 < \theta < 1/2)$. B_{θ} continues to measure the complementarity in intermediates between the sectoral TFPs but it leads to less extreme values given that the range of its exponent is now limited between -1 and 0.

Fig. A.1: Gains in aggregate productivity through competitive allocation of resources

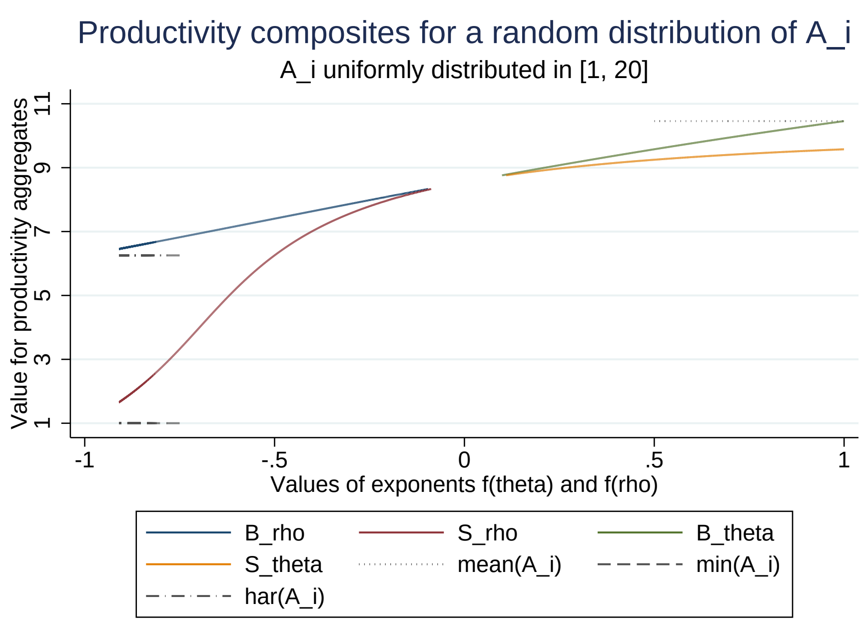


Figure A.1 illustrates the gains related to competitive allocation of resourcesin Equation A.7 with respect to the symmetric allocation of resources in Equation A.6. Values on the y-axis correspond to the results obtained for productivity aggregates S_θ, S_ρ, B_θ and B_ρ while values of x-axis correspond to values of their corresponding exponents. All possible combinations of composites S_ρ (red line) and S_θ (yellow line) for a uniform distribution of productivities in range [1, 20][31]. Limited cases are also plotted in this Figure and they correspond to minimum value of productivities ($min(A_i)$), the harmonic mean ($har(A_i)$) and the arithmetic mean of sectoral productivities ($mean(A_i)$). The lines drawn on the negative part of x-axis show the possible values for productivity aggregates in intermediates while those in the positive part of the x-axis are related to substituability in consumption.

Figure A.1 also shows the shift in aggregate productivity generated by competitive allocation as lines lying on the bottom of the Figure (red and yellow lines) are replaced by lines shifted upwards in the Figure (blue and green lines). The range of possible values for the composite based on intermediate consumption is substantially reduced and the lower bound is no longer the minimum level of productivity, but the harmonic mean of productivities. In consequence, all possible combinations of both productivity aggregates would have increased substantially. Nonetheless, we still observe that aggregate productivity will crucially depend on the distribution of productivities below their average value[32].

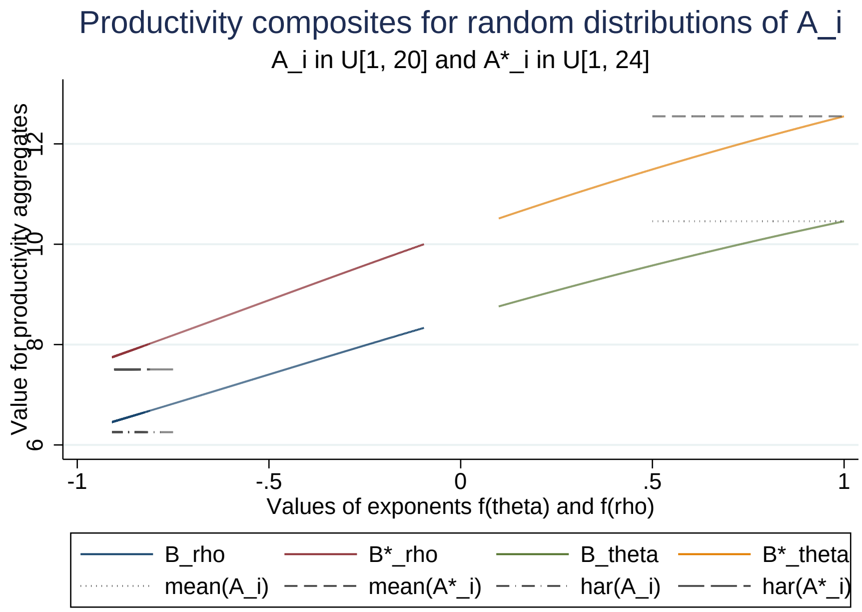
A second feature to be highlighted is the multiplier effect of these productivity aggregates. An increase of productivity by the same amount for all sectors is shown in Figure A.2. Limited values for sectoral productivites (mean and harmonic means) of both distributions are also plotted in the Figure. For all possible values of ρ and θ, we find that each of the productivity composites increases by the same amount. At the aggregate level, this linear increase of productivity leads to a more than proportional increase of productivity at the country level whose extent depends on the proportion of intermediate goods used in production (σ).

Starting for a random distribution of productivities, Figure A.3 shows the impact of left- and right-skewness on the range of possible values for the productivity composites. Results for three distributions are plotted here and their common characteristic is that their average value is identical. A_i^* is left-skewed and has a higher proportion of values below the average value which is also reflected by a lower value for the harmonic mean, the lower boundary for the term B_ρ. On the other hand, A_i^{**} is right-skewed and has a large proportion of observations above the average value. Despite being more concentrated on higher values we can observe that right-skewness has a similar effect

[31]Remember that the x-axis used in Figure A.1 are not values of θ and ρ but the exponents in A.8 and A.9. Nonetheless, given that these expressions are monotonic functions of values in θ and ρ, they help us in the illustration of comparisons.

[32]The shape of curves describing possible values for productivity composites is relatively flat and monotonic given that a uniform distribution of productivities is used for illustration.

Fig. A.2: Linear increase of productivity in the economy



as left-skewness with respect to a symmetric distribution, even though losses in aggregate productivity are less pronounced for right-skewed than in left-skewed distributions. Most productivity losses for A_i^{**} are concentrated in the term related to substitutability across intermediate goods (B_ρ^{**}). Indeed, the term related to consumption substitutability (B_θ^{**}) is only marginally different for the one observed for the symmetric distribution.

Figure A.4 shows the incidence of very low values on aggregate productivity by comparing two uniform distributions of productivities where aproximately 5% of observations of A_i^{**} are below the range of A_i. Low values lower the mean value of the distribution but their influence to this characteristic is rather limited while compared to changes observed for the composite of intermediate productivities (difference between blue and red lines). Assuming that the share of intermediates in production (σ) is one half, the incidence of the extreme observations implies losses at the country level which are equivalent to 14 % in the most favourable case but might attain 34% when substitutability across intermediate and final goods is the most constrained.

Fig. A.3: Left- and right-skewness of the distribution of productivities

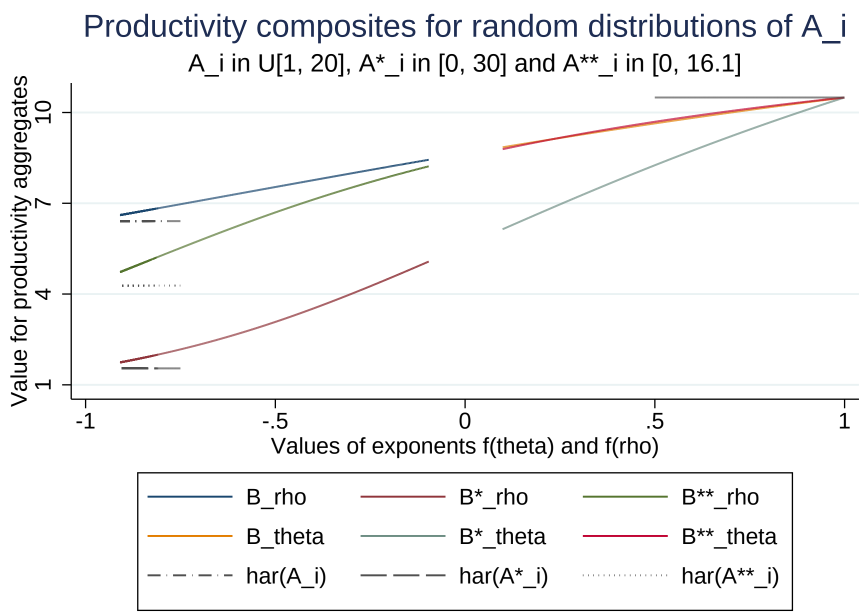
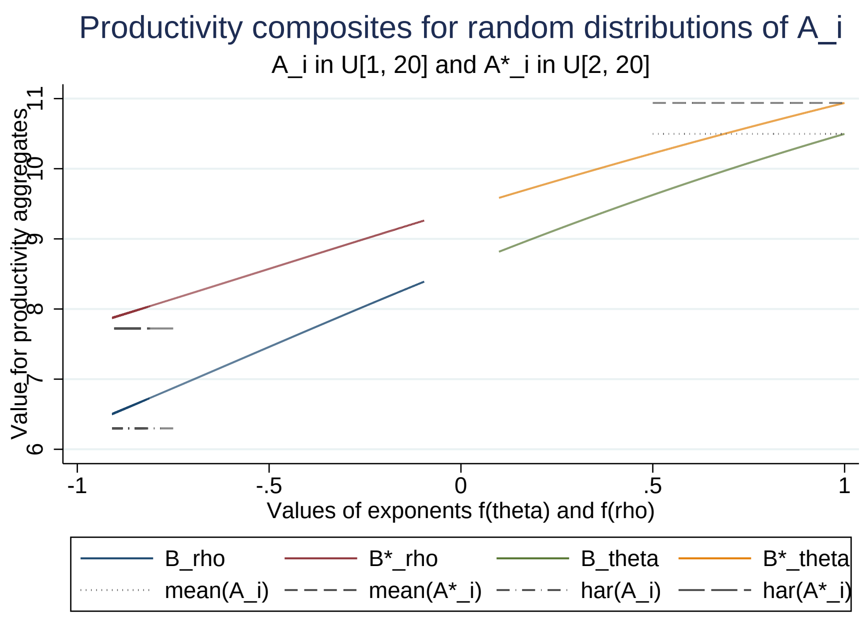
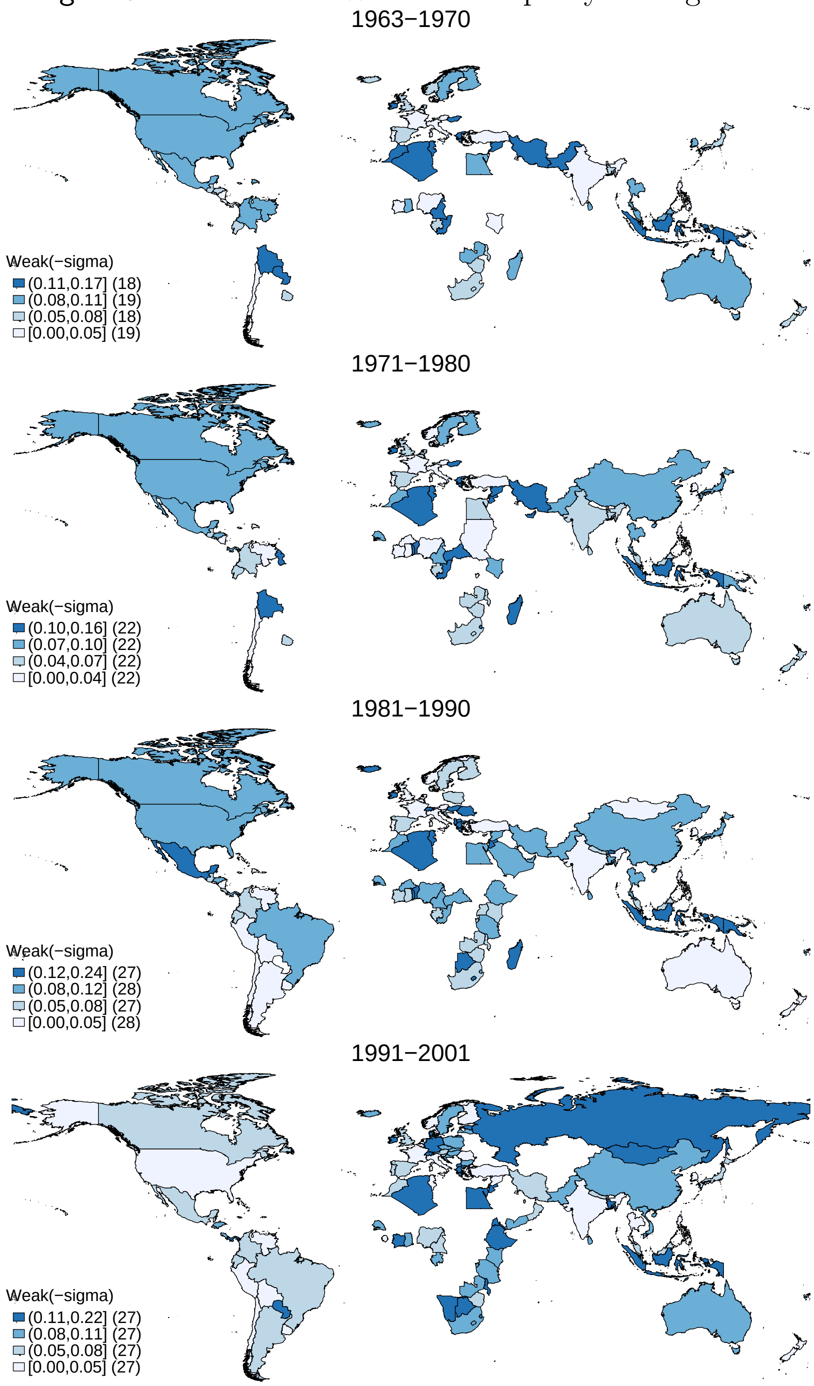


Fig. A.4: Fat tails in low values of the productivity distribution



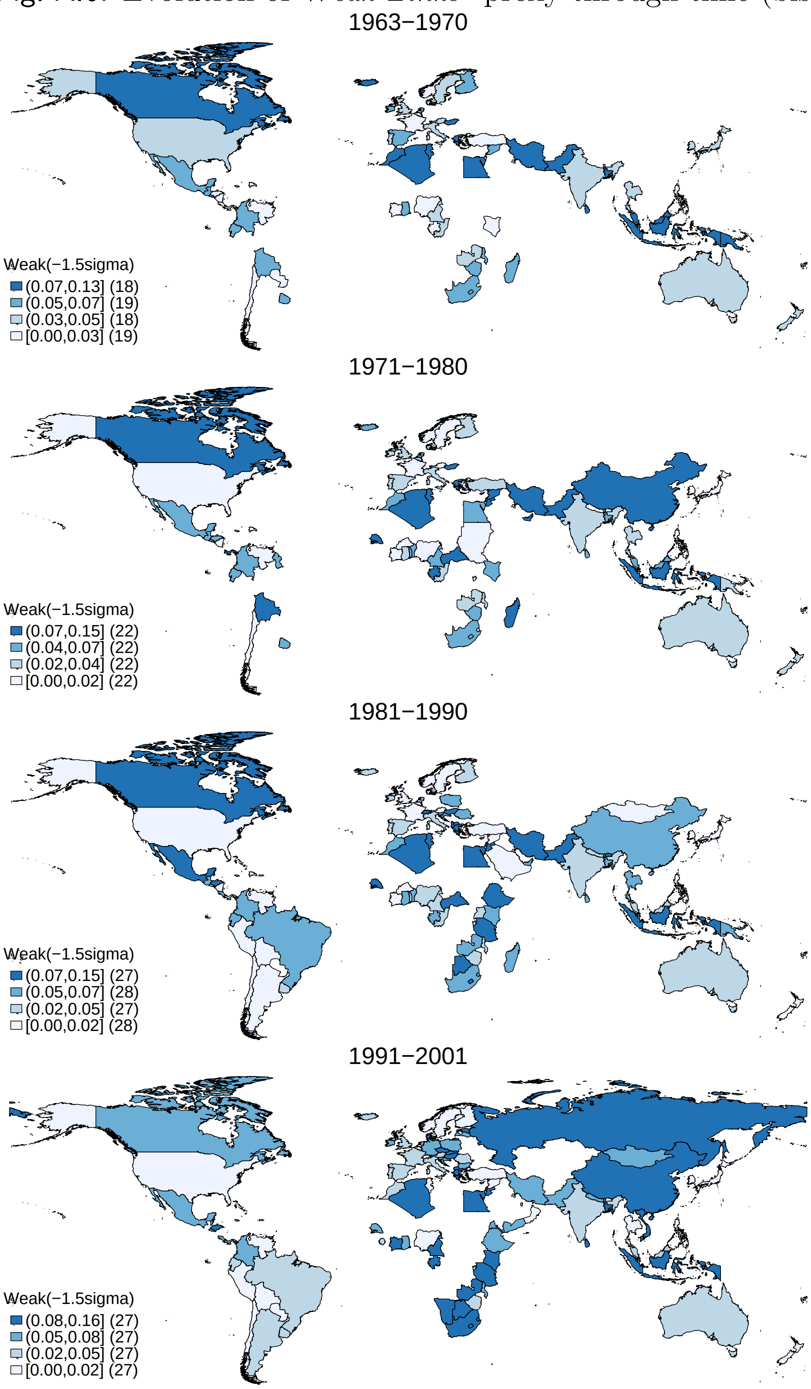
A.2 Supplementary Figures and Tables

Fig. A.5: Evolution of *Weak Links'* proxy through time



Note: The *Weak Links'* proxy used in this Figure is the probability of observing productivities lower than the mean productivity minus the standard deviation averaged over all observations for each country in each period considered above.

Fig. A.6: Evolution of *Weak Links'* proxy through time (bis)



Note: The *Weak Links'* proxy used in this Figure is the probability of observing productivities lower than the mean productivity minus 1.5 times the standard deviation averaged over all observations for each country in each period considered above.

Tab. A.1: Sample coverage

Albania (3), Algeria (28), Argentina (14), Australia (35), Austria (37), Bahamas (8), Bangladesh (28), Barbados (28), Belgium (35), Belize (2), Benin (7), Bhutan (1), Bolivia (31), Botswana (12), Brazil (5), Bulgaria (11), Burkina Faso (10), Burundi (17), Cameroon (25), Canada (39), Central African Republic (16), Chile (38), China (22), Colombia (38), Congo, Rep. (14), Costa Rica (21), Cote d'Ivoire (21), Croatia (3), Cyprus (27), Czech Republic (2), Denmark (29), Dominican Republic (23), Ecuador (37), Egypt, Arab Rep. (35), El Salvador (29), Eritrea (10), Estonia (2), Ethiopia (21), Fiji (25), Finland (38), France (31), Gabon (8), Gambia (8), Germany (3), Ghana (28), Greece (36), Guatemala (19), Guyana (1), Honduras (26), Hong Kong (29), Hungary (38), Iceland (29), India (39), Indonesia (32), Iran, Islamic Rep. (36), Ireland (38), Israel (39), Italy (34), Jamaica (28), Japan (39), Jordan (27), Kenya (39), Korea, Rep. (39), Kuwait (30), Latvia (9), Lesotho (4), Luxembourg (38), Macao, China (20), Macedonia, FYR (7), Madagascar (22), Malawi (32), Malaysia (33), Malta (39), Mauritius (19), Mexico (31), Mongolia (6), Morocco (24), Namibia (1), Nepal (9), Netherlands (38), New Zealand (34), Nicaragua (21), Nigeria (28), Norway (39), Oman (9), Pakistan (30), Panama (37), Papua New Guinea (27), Paraguay (6), Peru (14), Philippines (34), Poland (11), Portugal (38), Puerto Rico (16), Qatar (2), Romania (12), Russian Federation (9), Saudi Arabia (1), Senegal (24), Sierra Leone (1), Singapore (39), Slovak Republic (5), Slovenia (12), South Africa (30), Spain (38), Sri Lanka (28), Sudan (1), Swaziland (16), Sweden (38), Switzerland (11), Syrian Arab Republic (35), Tanzania (9), Thailand (18), Togo (10), Tonga (1), Trinidad and Tobago (32), Tunisia (28), Turkey (33), Uganda (1), United Arab Emirates (4), United Kingdom (34), United States (38), Uruguay (31), Venezuela, RB (34), Vietnam (1), Yemen, Rep. (4), Zambia (18), Zimbabwe (34).

In parenthesis, the number of observations per country.

Tab. A.2: Does measuring growth over different periods create a bias ?

$\Delta \ln GDP$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dummy	0.007	0.008	0.007	0.008	0.011	0.012*	0.011	0.012*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Δ *Weak* $(-\sigma)$	-0.046*				-0.040*			
	(0.025)				(0.024)			
Δ *Weak* (-1.5σ)		-0.087**				-0.085**		
		(0.035)				(0.034)		
$\Delta \ln P_{-\sigma}$			-0.053**				-0.046*	
			(0.027)				(0.026)	
$\Delta \ln P_{-1.5\sigma}$				-0.093**				-0.091**
				(0.037)				(0.036)
Constant	0.021***	0.021***	0.021***	0.021***	0.047***	0.048***	0.047***	0.048***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)	(0.007)	(0.007)	(0.007)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE					✓	✓	✓	✓
Observations	2689	2689	2689	2689	2689	2689	2689	2689
R^2	0.172	0.173	0.172	0.173	0.241	0.242	0.241	0.242

$\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. Δ *Weak* $(-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Dummy takes value 1 if annual growth rate is measured using non-consecutive observations over time. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tab. A.3: Does measuring growth over different periods create a bias ? (v2)

$\Delta \ln GDP$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(Time span)	0.004	0.005	0.004	0.005	0.006	0.007	0.006	0.007
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Δ *Weak* $(-\sigma)$	-0.046*				-0.040*			
	(0.025)				(0.024)			
Δ *Weak* (-1.5σ)		-0.086**				-0.085**		
		(0.035)				(0.034)		
$\Delta \ln P_{-\sigma}$			-0.053**				-0.046*	
			(0.027)				(0.026)	
$\Delta \ln P_{-1.5\sigma}$				-0.093**				-0.091**
				(0.037)				(0.036)
Constant	0.018***	0.017***	0.018***	0.017***	0.043***	0.043***	0.043***	0.043***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.008)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE					✓	✓	✓	✓
Observations	2689	2689	2689	2689	2689	2689	2689	2689
R^2	0.172	0.172	0.172	0.172	0.241	0.242	0.241	0.242

$\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. Δ *Weak* $(-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Time span is the number of years separating observations used to calculate the growth rate. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Tab. A.4: Growth regressions for selected years (1965, 1970, ..)

$\Delta \ln GDP$	(1)	(2)	(3)	(4)
Δ *Weak* $(-\sigma)$	-0.042*		-0.016	
	(0.024)		(0.023)	
Δ *Weak* (-1.5σ)		-0.085***		-0.052*
		(0.032)		(0.030)
Constant	0.020***	0.020***	0.036***	0.036***
	(0.001)	(0.001)	(0.003)	(0.003)
Country FE	✓	✓	✓	✓
Year FE			✓	✓
Observations	472	472	472	472
R^2	0.461	0.467	0.546	0.549

$\Delta \ln GDP$ stands for annual growth rate of per capita GDP at constant prices. Δ *Weak* $(-\sigma)$ is the change in the probability of observing *Weak Links* between two observations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

References

[1] Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar and Alireza Tahbaz-Salehi, 2012. The Network Origins of Aggregate Fluctuations. Econometrica 80(5), pp. 1977-2016, 09

[2] Barro, Robert J., 2000. Inequality and Growth in a Panel of Countries. Journal of Economic Growth, 5(1), pp. 5-32.

[3] Basu, Susanto, 1995. Intermediate Goods and Business Cycles: Implications for Productivity and Welfare. American Economic Review, 85(3), pp. 512-31.

[4] Ciccone, Antonio, 2002. Input chains and industrialization. Review of Economic Studies, 69(3), pp. 565-87.

[5] Dollar, David, 1992. Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976-1985. Economic Development and Cultural Change , 40(3), pp. 523-544.

[6] Frankel, Jeffrey A., and David Romer. 1999. Does Trade Cause Growth ? American Economic Review, 89(3), pp. 379- 99.

[7] Hirschman, Albert, 1958. The strategy of economic development. Yale University Press, New Haven.

[8] Islam, Nazrul, 1995. Growth Empirics: A Panel Data Approach. The Quarterly Journal of Economics, 110(4), pp. 1127-70.

[9] Jones, Charles I., 2011. Intermediate Goods and Weak Links in the Theory of Economic Development. American Economic Journal: Macroeconomics, 3(2), pp. 1-28.

[10] Kremer, Michael, 1993. The O-Ring Theory of Economic Development. The Quarterly Journal of Economics, 108(3), pp. 551-75.

[11] Levine, Ross, and David Renelt. 1992. A Sensitivity Analysis of Cross-Country Growth Regressions. American Economic Review, 84(4), pp. 942-63.

[12] Wacziarg, Romain, and Karen Welch, 2008. Trade Liberalization and Growth: New Evidence. World Bank Economic Review, 22(2), pp. 187-23.

Chapter III

Diversification and Weak Links

CHAPTER III
Diversification and Weak Links

Cristian Ugarte[*]

Department of Economics

University of Geneva, Switzerland

Abstract

An important literature has shown that the relationship between economic diversification and income per capita is non-monotonic (Imbs and Wacziarg, 2003 and Koren and Tenreyro, 2007). At early stages of development countries diversify as income increases and new economic opportunities emerge, but at later stages of development the production bundle becomes more concentrated as income rises. The aim of this paper is to explore the role played by *Weak Links* effects à la Jones (2011) in explaining the non-monotonic relationship between income per capita and economic diversification. To do so, we first construct a measure of the probability of observing *Weak Links* in a given country. Results show that economies where *Weak Links* are more likely to be observed tend to have a more concentrated production bundle. Moreover the inverted u-shape relationship between income per capita and economic diversification tends to be stronger in countries where *Weak Links* are more likely to be observed.

Keywords: Economic Diversification, Development, Weak Links.

[*]I am grateful to Celine Carrere, Olivier Cadot, Pramila Crivelli, Jorge Davalos, Jaime de Melo, Ndiame Diop, Jonathan Eaton, Liliana Foletti, Caroline Freund, Julien Gourdon, Alain McLaren, Marcelo Olarreaga, Silvio Tai and Laura Zoratto for very helpful discussions as well as all participants of the DSEC Informal Workshop at University of Geneva and the First International Conference of Development in Syria. All remaining errors are mine. E-mail: cristian.ugarte@unige.ch.

Contents

List of Tables

List of Figures

1 Introduction

Early stages of development are often accompanied by diversification of the production bundle as more economic opportunities become available. There is evidence, however, that the relationship between diversification and development is non-monotonic. Imbs and Wacziarg (2003) and Koren and Tenreyro (2007) show that if the two are positively correlated at low levels of development, once countries reach a certain income per capita threshold concentration of production increases with income levels. Cadot, Carrere and Strauss-Khan (2011) show that the u-shape relationship between economic concentration and income holds not only for production but also for export diversification.

A potential explanation for this empirical regularity is the tendency to diversify production (or investment opportunities) in the presence of incomplete financial markets at very low levels of development and the forces of comparative advantage that push towards concentration as financial markets develop (Gilles Saint Paul, 1992 and Acemoglu and Zillibotti, 1997). Faini (2004) suggests a similar explanation: at early stages of development as income rises the opportunities for risk diversification through sectorally diversified investment become stronger which initially leads to diversification. However as economies become richer, they also become economically and institutionally more stable, and this reduces business risks which reduces the incentives to diversify.

The presence of *Weak Links* à la Jones (2011) can have an impact on this relationship. Jones (2011) defines *Weak Links* building on the earlier work by Hirschman (1958) and Kremer (1993) that emphasize the role played by linkages and complementarities in economic development. Low productivity in one input sector for which there is little substitutability will act as a weak link in the production chain, hurting all downstream sectors and the overall development prospects of the country. Thus, the presence of *Weak Links* is likely to lead to a less diversified production bundle (ceteris paribus) as downstream sectors are hurt by higher input and associated factor prices due to the low productivity of this input sector for which there is little substitutability. On the other hand, in economies with a higher probability of observing a weak link, there are higher risks for investors which creates incentives for portfolio diversification to minimize the value at risk which would lead to a more diversified production bundle.

Therefore, the effect of *Weak Links* on economic diversification is ambiguous and remains an empirical matter. The aim of this paper is twofold. First, to explore the impact that the presence of *Weak Links* could have on the concentration of the production bundle. Second, to try to examine whether the presence of *Weak Links* could help explain the u-shaped relationship between income per capita and diversification. Thus, after creating a proxy for *Weak Links* that captures the probability that there exists a relatively low productive sector which is heavily used as an input by the rest of the economy, we estimate its impact on the degree of economic diversification and explore how the the relationship between economic diversification and income per capita varies across economies with high and low propensity of observing *Weak Links*[1].

Empirical results show that economies where *Weak Links* are more likely to be observed tend to have a more concentrated production bundle. Also the inverted u-shaped relationship between income per capita and economic diversification tends to be steeper for countries where *Weak Links* are more likely to be observed. However, it is important to notice that turning points (GDP levels) at which the slopes change sign are similar for economies with low and high levels of *Weak Links*.

The rest of the paper is organized as follows. Section 2 describes the empirical methodology and how the proxy for *Weak Links* is built. Section 3 presents the results and shows some robustness checks. Finally, section 4 concludes.

2 Empirical Methodology

Our starting point is Imbs and Wacziarg (2003) where the following relationship is estimated parametrically :

$$\text{Concentration}_{c,t} = f(\text{income}_{c,t}) + \sum_{c} \gamma_c D_c + \varepsilon_{c,t} \tag{1}$$

where $\text{Concentration}_{c,t}$ measures the lack of sectoral diversification[2] using different indices and along the value-added dimension for country c in year t, income is the GDP per capita at constant prices, noted $\text{GDPpc}_{c,t}$ hereafter, and $\varepsilon_{c,t}$ is an error term. The relationship, f, is estimated in an unbalanced panel of 98 countries over the period 1963-2001, and in order to control for unobserved heterogeneity country fixed effects (D_c) are included in all our estimations. Imbs and Wacziarg (2003) found a quadratic relationship between sectoral concentration and income per capita and this is our starting point.

We then explore how *Weak Links* affect the relationship between development and diversification. First, we simply add the proxy for *Weak Links* to Equation (1) :

$$\text{Concentration}_{c,t} = f(\text{income}_{c,t}) + \alpha\, \textit{Weak Links}_{c,t} + \sum_{c} \gamma_c D_c + \varepsilon_{c,t} \tag{2}$$

[1] This paper is closely related to the work by Olarreaga and Ugarte (2012) where authors use a *Weak Links'* measure to explain the tendency towards concentration of *Weak Links'* countries while using non-parametric and fourth order regressions to describe their development patterns. However, they do not estimate the impact of *Weak Links* on the well-established U-shaped pattern of development suggested by Imbs and Wacziarg(2003) as done here.

[2] Here we follow most of the literature and use indices of concentration rather than diversification.

Afterwards, we will explore non-monotonic effects of the proxy on the relationship between income and concentration, f, and it considers the impact of two versions of the proxy capturing the presence of *Weak Links*.

2.1 Measuring Diversification

We measure economic concentration using different indices to check for the robustness of the results. They will be calculated based on sectoral value-added for each country and year using the 28[3] sector disaggregation provided at the 3-digit level of the International Standard Industrial Classification of All Economic Activities (ISIC) rev. 3[4].

An important issue in the calculation of the concentration indices is non reported data. If data for some small sectors is missing and we instead consider it as a zero, this may increase the value of some indices like the Gini index but this will not affect the Herfindhal index. Thus given that the data is unbalanced within countries over time[5], especially for small sectors, we consider small sectors[6] as inactive if their information is missing and they are not included in the calculation of the concentration indices. The Herfindhal will not be affected by this rule, but the Gini might be downward biased. On the other hand, the Hefindhal is very sensitive to large sectors, whereas the Gini is more sensitive to what occurs in the middle of the distribution. So, these indices will capture differently changes towards diversification. Thus, to ensure the robustness of our results we also use the Theil index which puts a heavier weight to changes at the bottom of the distribution. The problem with this index is that at the bottom of the distribution we are not sure whether values close to zero are related to active or inactive sectors. Thus, we may be putting too much weight on noisy data by using a Theil index of concentration.

Thus our indices of diversification are the following: Gini, Herfindhal and normalized

[3]The number of sectors may vary per country and year due to misreport in the data or inactive sectors in a given economy. On average, we observe 21 active sectors per country and three quarters of countries included in the sample report data for 17 economic sectors. Only one every ten countries reports between 8 to 12 sectors on average.

[4]The source of the data is United Nations Industrial Development Organization (UNIDO) Industrial Statistics Database (INDSTAT). http://www.unido.org/.

[5]Thus, the number of sectors reporting data varies by country and year in our sample. The sample of developing countries used in the regressions covers 98 countries with an average of 18 country-year observations per country and one third of countries has at least 25 of 39 possible observations. Another third of countries in the sample has between 10 and 25 country-year observations and one third of countries have less than 10 observations over the period 1963-2001.

[6]Small sectors are those whose average share in total value-added over the whole period represents less than 2 percent. If the average share in total value-added exceeds 2 percent, the sector is no longer considered as small and we do not consider observations in that year in that country because of incomplete information.

Theil[7]. After ordering the value-added shares in increasing order, the Gini coefficient is calculated as follows :

$$\text{Gini} = 1 - \frac{1}{n_{c,t}} \sum_{i=1}^{n_{c,t}} \left(S_{i-1}^{c} + S_{i}^{c} \right) \tag{3}$$

where S_i^c is the cumulative share of value-added of sector i in country c, $n_{c,t}$ is the number of active sectors in country c at period t and $S_0 = 0$. The Gini index ranges between 0 and 1. When it takes the value zero in a homogeneously diversified economy where all sectors have an equal share of total value-added. At the other extreme, the economy is fully concentrated and all value-added is generated by a single sector.

The Herfindhal index ranges between $\frac{1}{n_{c,t}}$ and 1 and it also increases with the degree of concentration :

$$\text{Herf} = \sum_{i=1}^{n_{c,t}} (S_i)^2, \tag{4}$$

where S_i is the share of sector i in total value-added.

Finally, we use the normalized Theil index which is more sensitive to changes at the bottom of the distribution. Furthermore, it has similar range and properties than for the two previous indices :

$$\text{Theil}_{trad} = \frac{1}{n_{c,t}} \sum_{i=1}^{n_{c,t}} \left(\frac{S_i}{\overline{S}} \right) \times \log \left(\frac{S_i}{\overline{S}} \right) = \frac{1}{n_{c,t}\overline{S}} \left(\sum_{i=1}^{n_{c,t}} S_i \times \log(S_i) \right) - \log(\overline{S}) \tag{5}$$

$$\text{Theil}_{norm} = 1 - \exp(-\text{Theil}_{trad}), \tag{6}$$

where $\overline{S}$ is the average share of value-added, namely $1/n_{c,t}$.

2.2 A proxy for Weak Links

To estimate the impact of *Weak Links* on economic diversification as suggested in Equation 2, we use the proxy proposed in Chapter II of this Thesis which aims at capturing the occurrence of low productivity in intermediate goods' production.

[7] The normalization ensures that the range of the new index is $[0, 1]$ instead of $[0, \log n_{c,t}]$ for the traditional version of the Theil index.

Using residuals of labor productivity ($\mathrm{r.q}_{c,s,t}$) on country-year and sector-year fixed effects to approximate total factor productivities by sector[8] as shown in Equations 7-8, we run a kernel density estimation for the productivity distribution in intermediates faced by each country in each period[9].

$$\mathrm{q}_{c,s,t} = \frac{\text{Value added}_{c,s,t}}{\text{Labor}_{c,s,t}} = \sum_{c,t} \lambda_{c,t} + \sum_{s,t} \gamma_{s_t} + \varepsilon_{c,s,t} \tag{7}$$

$$\hat{\varepsilon}_{c,s,t} = \mathrm{r.q}_{c,s,t} \tag{8}$$

The proposed proxy for *Weak Links* measures the probability of observing abnormal low productivities using a relative threshold (λ standard deviation below the mean) by country and year, for example :

$$Weak_{c,t} = \text{Prob}\left[\widehat{\mathrm{r.q}} < mean_{c,t}(\widehat{\mathrm{r.q}}) - \lambda \times std_{c,t}(\widehat{\mathrm{r.q}})\right] \tag{9}$$

Here, we present the results for the proxy calculated using $\lambda = 1$ but we also show some results obtained using other values of λ[10]. For further details and characteristics of the proxy used, see the descriptive analysis of this measure in Chapter II.

In a framework where all economic sectors are considered producers of intermediate goods in the production of the final good, *Weak Links* should reflect the fact that relative low productive sectors will reduce the total productivity and implies that resources allocation is not optimal[11]. Less productive sectors will require more resources and this will harm the production of other productive sectors. If less productive sectors are not tradeable and incentives or possiblities of investment beyond these value chains are limited, this situation will lead to a higher concentration of the economy and thus, we expect that $\alpha > 0$ in Equation (2).

[8]The use of residual productivities is a major difference with respect to the results presented by Olarreaga and Ugarte (2012) where authors show that the *Weak Links*' proxy enables to distinguish between countries with an initial tendency to diversification in their patterns of development and countries with an initial tendency to concentration.

[9]The share in intermediate sales as well as an index of tradeability are used to weight the importance of each sector in the kernel density estimation.

[10]Several values for $\lambda \in [0.84, 2]$ were considered as we do not have a precise way of choosing this parameter.

[11]See Jones (2011).

3 Results

3.1 Data and Sample Description

Table A.1 lists the countries in the sample, as well as the number of observations available for the period 1963-2001. The unbalanced nature of our panel suggests that after controlling for country fixed effects our coefficients will mainly capture the within country variability of those countries with a large number of observations. The data on value-added comes from UNIDO's INDSTAT 2 and GDP per capita in constant prices is from the World Development Indicators. We create a correspondence between OECD input-output tables[12] and sectors in INDSTAT in order to establish the importance of each sector s as an intermediate good in each economy[13].

As a start, we replicate Imbs and Wacziarg (2003) to check that the U-shaped pattern of development is also valid in our sample which contains few more countries and a longer span in time. As our goal is to explore the relationship between economic concentration and development for developing countries, we focus on countries with per capita income below 15 thousands US$ and we exclude high-income countries. In fact, among the excluded countries we find some of the largest exporters of natural resources (oil) with significantly high levels of concentration and considering them in our sample would make our results less relevant for developing countries. In fact, these countries do not necessarily follow the usual pattern of diversification and given important rents of natural resources, they can shift their production structure more easily to specific sectors through investment.

[TABLES 1 AND 2 HERE]
[SUMMARY STATISTICS AND CROSS-CORRELATIONS' TABLE]

Tables 1 and 2 present summary statistics for the different measures of diversification. The correlation between measures is statistically significant at 1% level and positive in all cases. Table 1 also provides descriptive statistics of the probability of observing *Weak Links* in our sample. On average, the probability of observing low productivity in intermediates is around 7% for the threshold at one standard deviation with respect to the mean and it falls when the threshold is fixed at a lower value. The correlation associated to the measure, calculated using one standard deviation to the mean, is very low but significant and positive with all the concentration indices. The correlations using a second measure ($\lambda = 1.5$) are less significant and show much less correlation between the proxy and the diversification measures.

[12] http://www.oecd.org/sti/inputoutput/

[13] As only 24 countries report input-output tables, we average values of weights ω_s by region for all remaining countries. The regions defined are Latin America, Europe, Middle East & North Africa, Africa and Asia.

[FIGURES 1 AND 2 HERE]
[AVERAGE INDICES OF ECONOMIC CONCENTRATION ACROSS COUNTRIES (1963-2001)]
[AVERAGE VALUES OF WEAK LINKS' PROXY ACROSS COUNTRIES (1963-2001)]

Figure 1 provides a graphical representation of our sample and shows the substantial heterogeneity that exists across countries while measuring economic concentration. Concentration seems to be particularly relevant for African and Latin American countries, but even within each of these regions the variability of indices is significant. Figure 2 plots the average value of the proxy for *Weak Links* for each country in our sample. In this sense, Latin American economies seem to be less subject to observe relatively low productivities than other developing countries in other regions of the world. However, it can be observed that the higher probabilities of observing *Weak Links* are mainly associated with higher economic concentration[14] [15].

3.2 Results

First, we replicate Imbs and Wacziarg (2003) with our sample and found similar results to theirs. The u-shaped pattern between concentration and level of development is verified for all our measures of concentration and is robust to the inclusion of country and year fixed effects. These results are presented in the first 6 columns of Table 3. The last two columns of this Table show the results by Imbs and Wacziarg that are based on the same data source. Results obtained by Imbs and Wacziarg are very close to those observed in columns 1 and 3 of this Table, with a slightly increased sample size.

[TABLE 3 HERE]
[REPLICATING IMBS AND WACZIARG'S RESULTS]

Our first attempt is to include the proxy for *Weak Links* as a control variable in the regression as suggested by the specification in Equation 2. Table 4 shows the results obtained in this exercise. At this point, it seems that the proxy included does not capture any particular effect on concentration and estimates are rarely significant. This result could be partially explained by the low correlation observed in Table 2. However, we believe that these results are driven by a non-monotonic impact of *Weak Links* on concentration and Table 5 addresses this question by interacting the proxy with the income per capita and the square of it. All interactions and the proxy itself are highly

[14] The Russian Federation and China are two clear exceptions to this behavior.

[15] Further details and characteristics of the proxy used here can be found in Chapter II in Ugarte (2013).

significant in all regressions. The proxy has a positive impact on concentration and jointly with the effect of interactions we can argue that the presence of relatively less productive sectors in a country tend to be associated with higher levels of concentration at the sectoral level. Indeed, the coefficients of interactions of GDP and the proxy are significantly larger than those of GDP and its quadratic term. This points out a steeper u-shaped pattern for countries with higher propensity to observe *Weak Links*.

[TABLES 4 AND 5 HERE]
[INCLUDING WEAK LINKS' PROXY]
[INCLUDING WEAK LINKS AND INTERACTIONS WITH GDPpc]

Table 6 goes further in this exercise and splits the sample with respect to the mean probability of observing relatively low productive sectors[16]. We define a dummy variable taking the value 1 if the probability of observing low productivities is higher than the average probability in the sample. Using interactions terms with the dummy variable, we verify that the u-pattern exists for the two sub-samples. Except for the Herfindhal indices in countries with a high propensity to relatively low productivity in intermediates, all u-shaped patterns are confirmed by the in-sample test proposed by Lind and Mehlun (2010)[17]. Furthermore, we test the estimates between sub-samples and we find that the shapes of the u-shaped relationships are different for each group. In fact, countries with *Weak Links* show a steeper pattern which implies higher levels of concentration at earlier and later stages of development. Moreover, these countries with high probability of *Weak Links* remain more concentrated than other countries during their development as the estimate for the dummy variable is statistically significant in all regressions.

[TABLE 6 HERE]
[SPLITTING THE SAMPLE BETWEEN COUNTRIES WITH HIGH AND LOW PROPENSITY TO OBSERVE WEAK LINKS]

Figures 3.1-3.3 plot these patterns for the three indices of concentration in Table 6. The intersection of the two patterns drawn in these Figures is a potential explanation to the low statistical significance of the proxy in Table 4. These graphs show that poor countries with *Weak Links* show a lower correlation between diversification and income

[16]This distinction leads to two subsamples of almost similar sizes: 840 and 896 observations. Considering a splitting around the median value of the proxy would imply similar samples to those considered here.

[17]This test goes beyond the check of two necessary conditions, i.e. statistical significance of the coefficient of the quadratic term and the existence of an extremum point within the data, used to verify U-shaped relationships. Authors show that these two necessary conditions are not sufficient and they propose a test that checks the exact and necessary condition for the existence of a U-shaped relatioship in a finite sample.

level. In fact, these countries face important productivity bottlenecks that impede them of attaining higher levels of development. On the other hand, rich countries with *Weak Links* show a lower correlation between concentration and income level than rich countries with low probability of observing relatively low productive sectors.

[FIGURES 3.1, 3.2 AND 3.3 HERE]
[DIVERSIFICATION AND WEAK LINKS USING GINI, THEIL AND HERFINDHAL INDICES]

3.3 Robustness checks

A recurrent concern on previous results is related to the restricted sample used in the analysis as we focus exclusively on current developing countries and the relevance of *Weak Links* on economic concentration might be particular to this reduced sample of countries. Tables 7 and 8 extend the sample to all countries below the threshold of 15 thousands US$ of GDP per capita irrespectively of their current level of development and they show similar results to those observed in Tables 5 and 6 for the restricted sample. Thus, we conclude that *Weak Links* do not only influence the pattern of development of (current) developing countries but they have also played a role in the development of more advanced economies.

[TABLES 7 AND 8 HERE]
[WEAK LINKS AND INTERACTIONS WITH GDPpc INCLUDING OBSERVATIONS FOR HIGH-INCOME COUNTRIES]
[SPLITTING THE SAMPLE BETWEEN COUNTRIES WITH HIGH AND LOW PROPENSITY TO OBSERVE WEAK LINKS INCLUDING OBSERVATIONS FOR HIGH-INCOME COUNTRIES]

Moreover, the choice of $\lambda = 1$ is arbitrary as we lack from a clear criterion establishing the most suitable value for this parameter and a sensitivity analysis of our results using different values of λ is therefore appropriate. Tables 9 and 10 show the results obtained while running similar regressions to those in Table 5 but for values of λ equal to 0.84 and 1.5, respectively[18]. Tables 9 and 10 confirm that *Weak Links* are positively correlated with economic concentration and that countries with higher propensity to observe such effects tend to have a steeper pattern of development as the estimates for interaction terms between the proxies and GDP per capita increase the absolute value of the slope at both extremes of the relationship. It is also particularly important to notice that the magnitude of estimates increases for higher values of λ. Undoubtedly, this fact is related

[18]Several values for the parameter λ were evaluated but we believe that these two examples illustrate results obtained with a larger number of values.

to kind of productivities that is captured by the proxy which has a higher sensitivity to extremely low productivities when $\lambda = 1.5$. Indeed, extremely low productivities are those whose effects on economic concentration are the strongest.

[TABLES 9 AND 10 HERE]
[WEAK LINKS AND INTERACTIONS WITH GDPpc FOR DIFFERENT VALUES OF λ]

Given that the proxy for the presence of *Weak Links* is derived from a kernel density estimation which relies on labour productivities and their corresponding weights as explained in previous section, measurement error on the estimation of the proxy is likely to arise. To address a potential measurement bias, we implement a correction suggested by Fuller (1986) and already used by Gawande et al. (2012) among others. This correction identifies values of the proxy for *Weak Links* that are estimated with low accuracy and in order to reduce their influence in the regression, the correction brings those values closer to the average value for the proxy in the sample.

We consider an additive measurement error for $Weak = P_i = \hat{P}_i + e_i$ where $\hat{P}_i$ is the true value of the probability and e_i is the error in its estimation. The variance of the error in each estimation (σ_i^2) is known since we estimated P_i using a kernel density of $1'000$ points and its expression is given by :

$$\sigma_i^2 = var[P_i] = var\left[\sum_k \mathbb{1}(k) \times p(k)\right] = P_i(1 - P_i) \times \sum_k [p(k)]^2 \tag{10}$$

where $p(k)$ is the probability of each point k in the kernel density estimation and $\mathbb{1}(k)$ is a dichotomous variable coding the probability of each point k of being a *weak link*.

Using the mean probability of observing *Weak Links* in the sample ($\bar{P}$), the sample variance of P_i (σ_E^2) and the mean variance of P_i ($\bar{\sigma}_u^2$), we apply the following transformation to the estimated probabilities :

$$\tilde{P}_i = \frac{\sigma_i^2}{\sigma_E^2 - \bar{\sigma}_u^2}\bar{P} + \left(1 - \frac{\sigma_i^2}{\sigma_E^2 - \bar{\sigma}_u^2}\right)(P_i - \bar{P}). \tag{11}$$

All values of the proxy which are estimated with low accuracy (high uncertainty) will be brought closer to the mean probability of observing *Weak Links* in the sample while values of the proxy which are accurately estimated will not be affected by the transformation. Changes of the proxy due to this correction might be observed in Figure 4. Points out of the diagonal are those that were not precisely estimated and the correction narrows them from the mean value of our proxy.

[FIGURE 4 HERE]
[CORRECTION OF THE WEAK LINKS' PROXY FOR MEASUREMENT ERROR]

Finally, Table 11 reports the estimates obtained once this transformation has been applied to our proxy $Weak$. Results are in line with those previously presented in Table 5 and so, the measurement error correction has not changed estimates significantly.

[TABLE 11 HERE]
[APPLYING AN ERROR CORRECTION IN THE ESTIMATION OF THE PROXY FOR WEAK LINKS]

The proxy used here is meant to capture the underlying productivity distribution that is faced by each country in terms of intermediate inputs. Therefore, it is worth evaluating whether the double weights considered in this exercise are determinant to understand the impact on economic concentration. Table 12 uses an alternative proxy for *Weak Links* which ignores the share of each sector in intermediates' sales and their level of tradeability. In fact, this alternative measure considers that all sectors have an equal importance as an input for other sectors in the economy and the interactions of the proxy with the GDP per capita as well as the proxy itself are no longer statistically significant as in previous versions of these regressions. We conclude that weights used in the kernel density estimation of productivity residuals are relevant and they play an important role in the explanation of the pattern of development followed by countries. These results are in line with recent findings in Acemoglu et al. (2012) which show that the structure and interactions between sectors of the economy affect the performance observed at the country level[19].

[TABLE 12 HERE]
[CALCULATING WEAK LINKS' PROXY WITHOUT SPECIFIC WEIGHTS FOR SECTORS]

4 Conclusions

This paper aims at measuring the impact of low productivity in intermediates' sectors on the u-shaped pattern of development described in Imbs and Wacziarg (2003). To quantify these effects, we use data on value-added and employment for manufacturing sectors in almost 100 developing countries for the period 1963-2001 available from UNIDO's INDSTAT 2. We combine this information with information coming from input-output tables in order to take into account the importance of each sector as an

[19] A similar conclusion has been raised in Chapter II in Ugarte (2013).

input for each economy and also its degree of substitutability by imported goods. The measure proposed gives a propensity to observe relatively low productivities in intermediate goods based on the distribution of productivities observed in each country and each year.

Our results show that *Weak Links* have a non-monotonic relation with economic concentration as *Weak Links* not only shift the u-shaped pattern of development toward more concentrated production bundles but they also affect the shape of the pattern. The different versions of the proxy proposed here have a significant and positive impact on indices of economic concentration. The magnitude of the effect increases as the proxy becomes more sensitive to extremely low productivities which unveils the severity of the costs related to these extreme values in terms of development[20]. Regarding the slope of the u-shaped relationship, it becomes steeper for low and high levels of development for countries with higher propensity to observe *Weak Links* and this points out that the benefits of diversification into new areas are lower for poor countries with high probability of observing *Weak Links*.

The previous conclusions are consistent to a number of robustness checks which include the definition of the proxy variable, the enlargement of the sample to developed countries and the treatment of potential measurement bias in the estimation of the proxy. We also provide evidence showing that the structure and sector linkages of each economy play a relevant role as the forementioned effects are only significant when inputs' need and tradeability indices are considered in the calculation of the proxy.

Finally, the evidence provided in this paper is highly relevant in the formulation of industrial and economic policies in developing countries. The benefits of production innovation are well known and the discovery and launch of new products has become a goal in the search of prosperity in most of countries. However, our study suggests that the correlation between diversification and income level is lower in the presence of *Weak Links* and that the benefits expected of diversification can be mitigated if countries do not address in a coherent and comprehensive way the production bottlenecks that lower their competitiveness.

[20] Jones (2011) shows that in case of low substituability, resources need to be deviated from other sectors to the low productive sector which increases the capital per worker in the *Weak Link* sector generating more concentration. In case of severe deviations of productivity, the displacement is larger than proportional which generates larger effect of this kind of deviations on economic concentration.

Tables and Figures

Fig. 1: Average indices of economic concentration across countries (1963-2001)

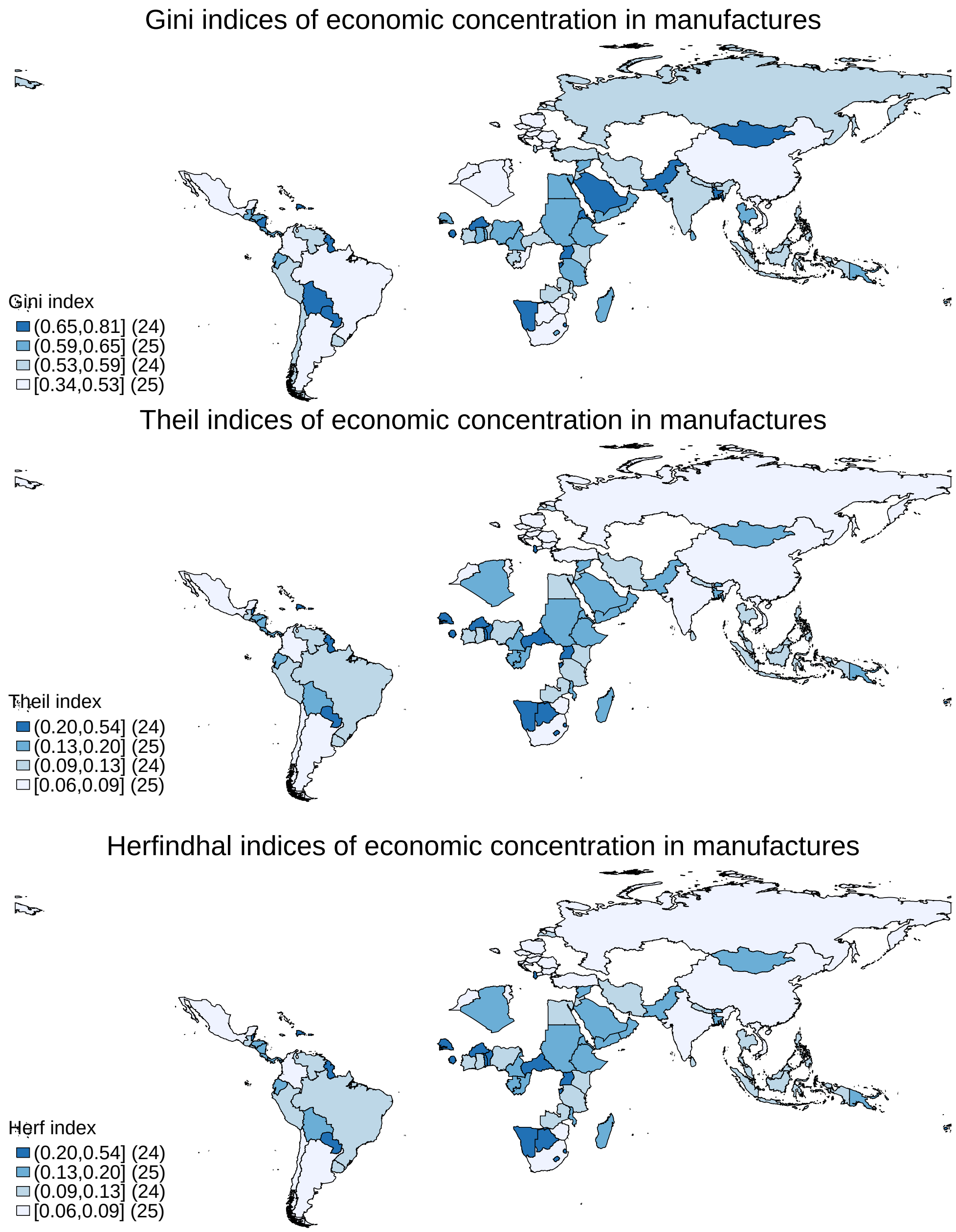
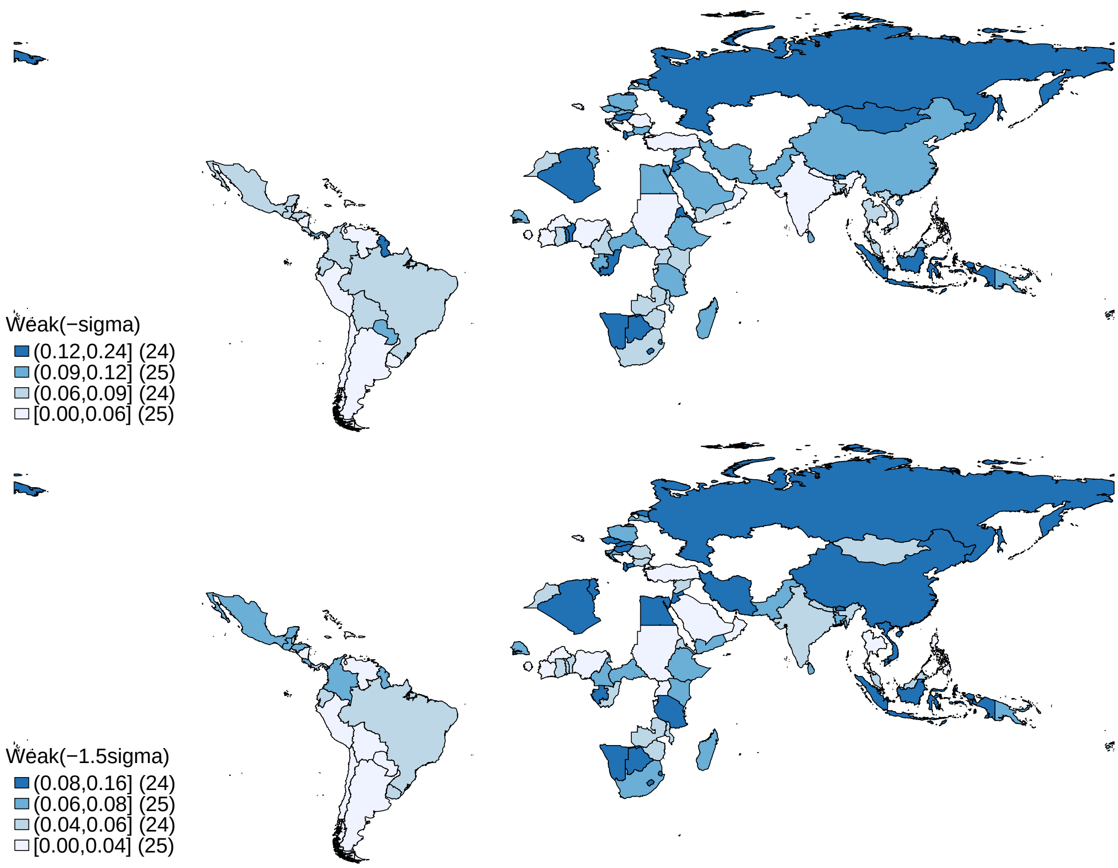


Fig. 2: Average value of *Weak Links'* proxy across countries



Note: The *Weak Links'* proxy used in the first panel of this Figure is the probability of observing productivities lower than the mean productivity minus the standard deviation averaged over all observations for each country in the period 1963-2001. The second panel uses the probability of observing productivities lower than the mean productivity minus 1.5 times the standard deviation as proxy.

Fig. 3.1: Diversification and *Weak Links* using Gini indices

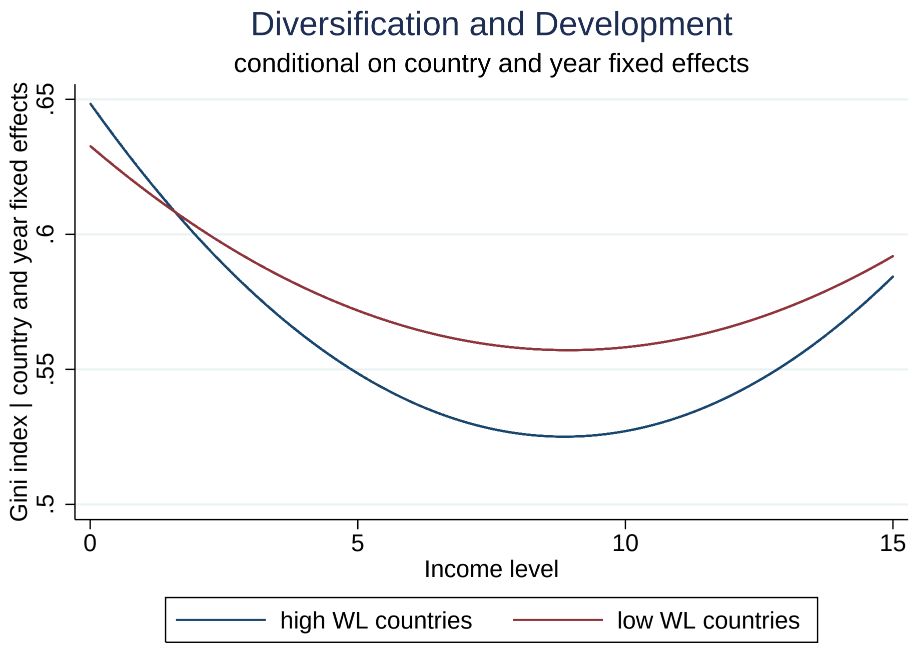


Fig. 3.2: Diversification and *Weak Links* using Theil indices

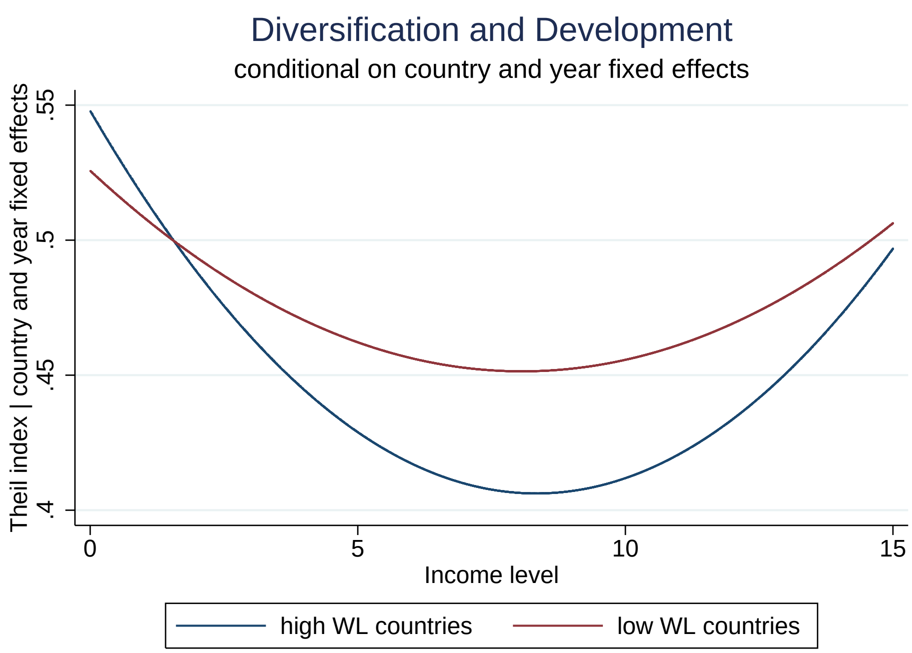


Fig. 3.3: Diversification and *Weak Links* using Herfindhal indices

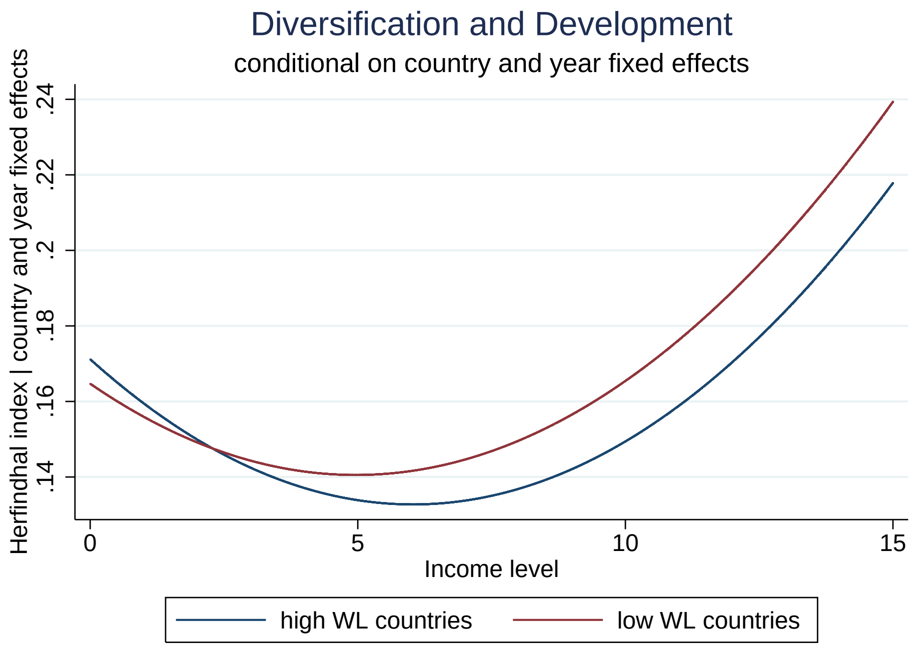
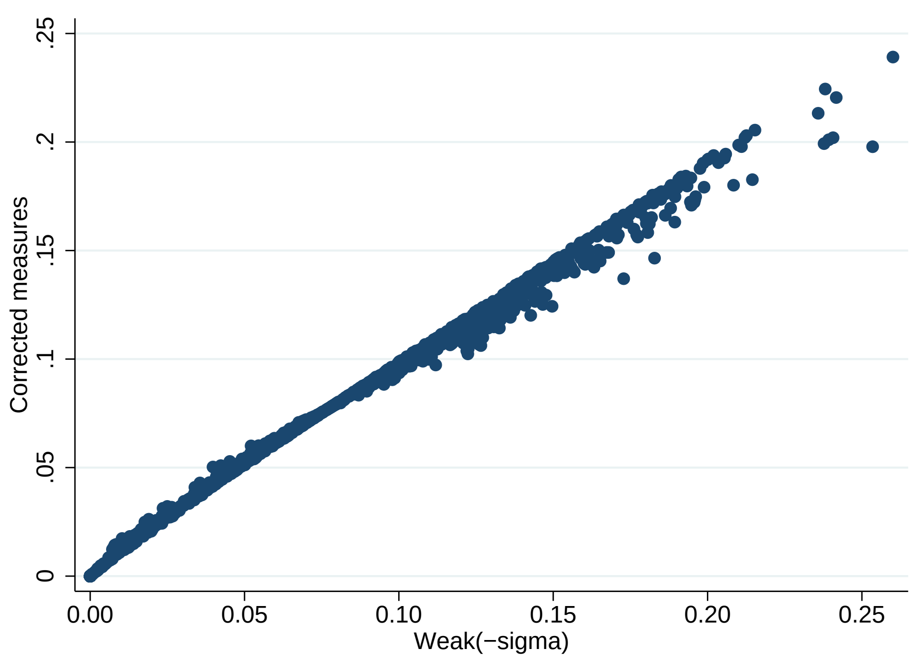


Fig. 4: Correction of the *Weak Links'* proxy for measurement error



Tab. 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Gini	0.583	0.093	0.293	0.822	1736
Theil	0.466	0.127	0.141	0.807	1736
Herf	0.138	0.077	0.055	0.763	1736
year			1963	2001	1736
GDP_{pc}	2.056	2.64	0.098	14.961	1736
Weak(-σ)	0.079	0.053	0	0.26	1736
Weak(-1.5σ)	0.054	0.043	0	0.236	1736

Tab. 2: Cross-correlation table

	$Gini_{vadd}$	$Herf_{vadd}$	$Theil_{vadd}$	*Weak(-σ)*	*Weak(-1.5σ)*
$Gini_{vadd}$	1.000				
$Herf_{vadd}$	0.712	1.000			
$Theil_{vadd}$	0.990	0.762	1.000		
Weak(-σ)	0.105	0.227	0.109	1.000	
Weak(-1.5σ)	0.003	0.022	0.003	0.718	1.000

Tab. 3: Replicating Imbs & Wacziarg

	Enlarged sample						Imbs & Wacziarg	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Gini	Theil	Herf	Gini	Theil	Herf	Gini	Herf
GDP_{pc}	-0.02377395***	-0.02912357***	-0.00636419**	-0.02288371***	-0.02690430***	-0.01140106***	-0.0161***	-0.0073***
	(0.00472136)	(0.00627614)	(0.00295271)	(0.00493388)	(0.00675760)	(0.00336880)		
GDP^2_{pc}	0.00135644***	0.00176318***	0.00085858***	0.00129549***	0.00164023***	0.00104703***	0.0009***	0.0003***
	(0.00022059)	(0.00030355)	(0.00019792)	(0.00021794)	(0.00030201)	(0.00018897)		
Constant	0.61648874***	0.50625585***	0.14164328***	0.64144721***	0.53796984***	0.16815732***	0.6126***	0.1308***
	(0.00761860)	(0.01005321)	(0.00446615)	(0.01227008)	(0.01663880)	(0.00761212)		
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓		
Observations	1736	1736	1736	1736	1736	1736	1493	1493
R^2	0.737	0.743	0.787	0.746	0.752	0.799	0.388	0.210

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Columns 7 and 8 report estimates from Table 4 of Imbs and Wacziarg (2003). Standard errors for columns 7 and 8 are not available in the original table.

Tab. 4: Including *Weak Links'* proxy

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
GDP_{pc}	-0.02373043***	-0.02907932***	-0.00619943**	-0.02291002***	-0.02695041***	-0.01129767***
	(0.00471905)	(0.00627596)	(0.00297832)	(0.00493457)	(0.00675955)	(0.00339743)
GDP^2_{pc}	0.00135985***	0.00176665***	0.00087149***	0.00129372***	0.00163713***	0.00105399***
	(0.00022162)	(0.00030464)	(0.00020078)	(0.00021808)	(0.00030175)	(0.00019061)
$Weak$	0.01427989	0.01452110	0.05406491**	-0.00831241	-0.01457194	0.03267029
	(0.03429266)	(0.04603367)	(0.02559996)	(0.03536155)	(0.04727783)	(0.02496320)
Constant	0.61523446***	0.50498039***	0.13689449***	0.64226889***	0.53941027***	0.16492789***
	(0.00802609)	(0.01070079)	(0.00508986)	(0.01276397)	(0.01736919)	(0.00807369)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	1736	1736	1736	1736	1736	1736
R^2	0.737	0.743	0.787	0.746	0.752	0.799

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 5: Including *Weak Links* and interactions with GDP_{pc}

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
GDP_{pc}	-0.01596905***	-0.01776465***	-0.00062130	-0.01554833***	-0.01595145**	-0.00645754
	(0.00500622)	(0.00664429)	(0.00349068)	(0.00526488)	(0.00715242)	(0.00397275)
GDP^2_{pc}	0.00081943***	0.00097389***	0.00051098**	0.00080345***	0.00089909**	0.00075974***
	(0.00025839)	(0.00035359)	(0.00025097)	(0.00025784)	(0.00035445)	(0.00024310)
$GDP_{pc} \times Weak$	-0.10212176***	-0.14917758***	-0.07168348***	-0.08844753***	-0.13255471***	-0.05608320**
	(0.03031444)	(0.04028246)	(0.02605525)	(0.03042558)	(0.04053095)	(0.02451866)
$GDP^2_{pc} \times Weak$	0.00578330**	0.00859906***	0.00320636	0.00493082**	0.00755247**	0.00229926
	(0.00245995)	(0.00324043)	(0.00236778)	(0.00247772)	(0.00325247)	(0.00212753)
$Weak$	0.17659829***	0.24983078***	0.17818991***	0.13439710**	0.19737046**	0.13299118***
	(0.06241995)	(0.08331666)	(0.04320797)	(0.06456597)	(0.08633284)	(0.04362501)
Constant	0.60308459***	0.48733637***	0.12777644***	0.63039291***	0.52174385***	0.15672692***
	(0.00879317)	(0.01169339)	(0.00570947)	(0.01343793)	(0.01821671)	(0.00854152)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	1736	1736	1736	1736	1736	1736
R^2	0.739	0.746	0.790	0.748	0.754	0.801

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 6: Splitting the sample between countries with high and low propensity to observe *Weak Links*

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
$GDP_{pc} \times (1 - dummy)$	-0.01846875***	-0.02172418***	-0.00447595	-0.01692320***	-0.01841562***	-0.00973704***
	(0.00482149)	(0.00630202)	(0.00307296)	(0.00510932)	(0.00684751)	(0.00355179)
$GDP^2_{pc} \times (1 - dummy)$	0.00102257***	0.00129539***	0.00076960***	0.00094704***	0.00114140***	0.00098108***
	(0.00023562)	(0.00031743)	(0.00021971)	(0.00023704)	(0.00032038)	(0.00021110)
$GDP_{pc} \times dummy$	-0.03044939***	-0.03842816***	-0.00893007***	-0.02786277***	-0.03400666***	-0.01275108***
	(0.00490583)	(0.00660571)	(0.00326201)	(0.00505048)	(0.00694436)	(0.00350573)
$GDP^2_{pc} \times dummy$	0.00170920***	0.00225826***	0.00092933***	0.00157262***	0.00204036***	0.00105753***
	(0.00025336)	(0.00034600)	(0.00022150)	(0.00024956)	(0.00034116)	(0.00020532)
dummy	0.01941248***	0.02640268***	0.01085002***	0.01577773***	0.02219688***	0.00647885*
	(0.00546766)	(0.00720448)	(0.00361164)	(0.00567782)	(0.00747150)	(0.00361466)
Constant	0.60721349***	0.49363955***	0.13665618***	0.63273165***	0.52571131***	0.16469586***
	(0.00777593)	(0.01020568)	(0.00475674)	(0.01279258)	(0.01720419)	(0.00787342)
Tests						
(1) $GDP_{pc} \times (1 - dummy) = GDP_{pc} \times dummy$	Reject	Reject	Reject	Reject	Reject	Accept
(2) $GDP^2_{pc} \times (1 - dummy) = GDP^2_{pc} \times dummy$	Reject	Reject	Accept	Reject	Reject	Accept
(1) + (2) Joint	Reject	Reject	Reject	Reject	Reject	Reject
Joint + Dummy	Reject	Reject	Reject	Reject	Reject	Accept
Lind and Mehlun test null hypothesis: relationship is not U-shaped						
p-value if dummy = 1	0.000	0.000	0.130	0.022	0.064	0.325
p-value if dummy = 0	0.013	0.040	0.046	0.000	0.002	0.000
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	1736	1736	1736	1736	1736	1736
R^2	0.740	0.746	0.788	0.749	0.755	0.800

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 7: *Weak Links* and interactions with GDP_{pc} including observations for high-income countries

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
GDP_{pc}	-0.00739608***	-0.00847311**	-0.00025508	-0.00386885	-0.00294378	-0.00165825
	(0.00257892)	(0.00349163)	(0.00179319)	(0.00287739)	(0.00389644)	(0.00201878)
GDP^2_{pc}	0.00060641***	0.00077001***	0.00029072**	0.00056092***	0.00067138***	0.00040404***
	(0.00014940)	(0.00020240)	(0.00012842)	(0.00015997)	(0.00021503)	(0.00013184)
$GDP_{pc} \times Weak$	-0.08326787***	-0.10454143***	-0.03989364**	-0.07337226***	-0.09324904***	-0.02623644
	(0.02136159)	(0.02821932)	(0.01832786)	(0.02228772)	(0.02961030)	(0.01821699)
$GDP^2_{pc} \times Weak$	0.00546106***	0.00615571***	0.00016425	0.00479382***	0.00541970**	-0.00069395
	(0.00164953)	(0.00221137)	(0.00168233)	(0.00171381)	(0.00229895)	(0.00164592)
$Weak$	0.15401741***	0.20205005***	0.14508101***	0.11761981**	0.15675805**	0.10251203***
	(0.05672508)	(0.07553344)	(0.03858527)	(0.05873728)	(0.07845686)	(0.03940641)
Constant	0.56860677***	0.44411933***	0.11661969***	0.59178092***	0.47268943***	0.13621468***
	(0.00691473)	(0.00936243)	(0.00430435)	(0.00968676)	(0.01308307)	(0.00598131)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	2315	2315	2315	2315	2315	2315
R^2	0.774	0.778	0.792	0.781	0.785	0.801

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 8: Splitting the sample between countries with high and low propensity to observe *Weak Links* (including high-income countries' observations)

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
$\text{GDP}_{pc} \times (1 - dummy)$	-0.00969487***	-0.01169473***	-0.00269794*	-0.00552563**	-0.00534285	-0.00380549**
	(0.00234215)	(0.00314323)	(0.00147070)	(0.00263963)	(0.00355013)	(0.00169916)
$\text{GDP}^2_{pc} \times (1 - dummy)$	0.00077874***	0.00098358***	0.00036032***	0.00069641***	0.00083825***	0.00045126***
	(0.00012635)	(0.00016809)	(0.00009296)	(0.00013401)	(0.00017687)	(0.00009583)
$\text{GDP}_{pc} \times dummy$	-0.01760995***	-0.02107915***	-0.00468192***	-0.01280415***	-0.01416495***	-0.00437594**
	(0.00270513)	(0.00364246)	(0.00174972)	(0.00285535)	(0.00389141)	(0.00187646)
$\text{GDP}^2_{pc} \times dummy$	0.00125667***	0.00150847***	0.00035185***	0.00113804***	0.00133438***	0.00036083***
	(0.00015929)	(0.00021176)	(0.00012258)	(0.00015775)	(0.00021081)	(0.00012017)
dummy	0.01318589***	0.01706087***	0.00848625***	0.01065370**	0.01430621**	0.00421002
	(0.00494008)	(0.00653304)	(0.00317268)	(0.00511248)	(0.00675516)	(0.00323919)
Constant	0.57349355***	0.45059474***	0.12375269***	0.59496775***	0.47669419***	0.14220022***
	(0.00580729)	(0.00785169)	(0.00338233)	(0.00898623)	(0.01207647)	(0.00529581)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	2315	2315	2315	2315	2315	2315
R^2	0.774	0.778	0.789	0.781	0.785	0.798

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 9: Including *Weak Links* and interactions with GDP_{pc} for $\lambda = 0.84$

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
GDP_{pc}	-0.01835059***	-0.02114052***	-0.00147412	-0.01807680***	-0.01949275**	-0.00716691
	(0.00519027)	(0.00696095)	(0.00385481)	(0.00603791)	(0.00810428)	(0.00444525)
GDP^2_{pc}	0.00102924***	0.00127260***	0.00061959**	0.00101628***	0.00119842**	0.00085660***
	(0.00033541)	(0.00044983)	(0.00024911)	(0.00035741)	(0.00047973)	(0.00026314)
$\text{GDP}_{pc} \times Weak$	-0.06343881**	-0.09329271***	-0.05381244***	-0.05282451**	-0.08079581**	-0.04262106**
	(0.02497069)	(0.03348954)	(0.01854573)	(0.02515622)	(0.03376550)	(0.01852059)
$\text{GDP}^2_{pc} \times Weak$	0.00293250	0.00441855	0.00203887	0.00233192	0.00373096	0.00142299
	(0.00214171)	(0.00287236)	(0.00159065)	(0.00215548)	(0.00289316)	(0.00158692)
Weak	0.09597962**	0.13717384**	0.15058674***	0.05290787	0.08348462	0.10906409***
	(0.04469425)	(0.05994187)	(0.03319441)	(0.04542952)	(0.06097699)	(0.03344626)
Constant	0.60866991***	0.49510509***	0.12797007***	0.63767178***	0.53172899***	0.15688751***
	(0.00842826)	(0.01130359)	(0.00625967)	(0.01306774)	(0.01753996)	(0.00962078)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	1736	1736	1736	1736	1736	1736
R^2	0.739	0.745	0.790	0.748	0.754	0.801

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 10: Including *Weak Links* and interactions with GDP_{pc} for $\lambda = 1.5$

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
GDP_{pc}	-0.01507452***	-0.01652590**	-0.00262226	-0.01555991***	-0.01606076**	-0.00885802**
	(0.00499475)	(0.00669612)	(0.00374129)	(0.00579901)	(0.00778160)	(0.00429238)
GDP^2_{pc}	0.00079780***	0.00095279**	0.00062686***	0.00082064***	0.00093713**	0.00089772***
	(0.00028824)	(0.00038642)	(0.00021590)	(0.00030911)	(0.00041479)	(0.00022880)
$GDP_{pc} \times Weak$	-0.21053854***	-0.30533008***	-0.07631638**	-0.18171257***	-0.26804132***	-0.04403065
	(0.05132846)	(0.06881255)	(0.03844733)	(0.05157464)	(0.06920714)	(0.03817512)
$GDP^2_{pc} \times Weak$	0.01573236***	0.02279024***	0.00259780	0.01394908***	0.02039829***	0.00046119
	(0.00496854)	(0.00666098)	(0.00372166)	(0.00499310)	(0.00670016)	(0.00369585)
$Weak$	0.35134239***	0.50405905***	0.09369936*	0.30859776***	0.44879512***	0.04829402
	(0.07158710)	(0.09597192)	(0.05362196)	(0.07197776)	(0.09658573)	(0.05327734)
Constant	0.59869964***	0.48080618***	0.13736928***	0.62550931***	0.51480450***	0.16548029***
	(0.00808658)	(0.01084112)	(0.00605721)	(0.01253978)	(0.01682692)	(0.00928184)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	1736	1736	1736	1736	1736	1736
R^2	0.740	0.747	0.788	0.749	0.755	0.800

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 11: Applying an error correction in the estimation of the proxy for *Weak Links*

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
GDP_{pc}	-0.01569604***	-0.01736894**	-0.00032094	-0.01540635**	-0.01573874*	-0.00622037
	(0.00514217)	(0.00689605)	(0.00382393)	(0.00597885)	(0.00802447)	(0.00440332)
GDP^2_{pc}	0.00081984***	0.00097287**	0.00050864**	0.00080913**	0.00090574**	0.00075722***
	(0.00031467)	(0.00042199)	(0.00023400)	(0.00033668)	(0.00045187)	(0.00024796)
$GDP_{pc} \times Weak_{error}$	-0.11068218***	-0.16166060***	-0.07840537***	-0.09599092***	-0.14375569***	-0.06256831***
	(0.03233454)	(0.04336314)	(0.02404531)	(0.03256273)	(0.04370389)	(0.02398194)
$GDP^2_{pc} \times Weak_{error}$	0.00606408**	0.00906092**	0.00348934*	0.00515128*	0.00793370**	0.00257219
	(0.00275104)	(0.00368936)	(0.00204579)	(0.00276931)	(0.00371681)	(0.00203955)
$Weak_{error}$	0.18767599***	0.26632577***	0.19946023***	0.14064550**	0.20780613***	0.15123755***
	(0.05683963)	(0.07622638)	(0.04226832)	(0.05767917)	(0.07741378)	(0.04247979)
Constant	0.60278317***	0.48683799***	0.12648013***	0.63053207***	0.52181455***	0.15558334***
	(0.00847802)	(0.01136969)	(0.00630461)	(0.01299837)	(0.01744569)	(0.00957309)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	1736	1736	1736	1736	1736	1736
R^2	0.739	0.746	0.790	0.748	0.754	0.801

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Tab. 12: Calculating *Weak Links'* proxy without specific weights ω for sectors

	(1) Gini	(2) Theil	(3) Herf	(4) Gini	(5) Theil	(6) Herf
GDP_{pc}	-0.02038751***	-0.02398992***	-0.00280733	-0.02055401***	-0.02303310***	-0.00866676**
	(0.00479800)	(0.00637162)	(0.00328555)	(0.00500278)	(0.00681051)	(0.00370235)
GDP^2_{pc}	0.00111718***	0.00139611***	0.00066721***	0.00113754***	0.00137310***	0.00091234***
	(0.00024214)	(0.00033380)	(0.00023586)	(0.00023967)	(0.00033153)	(0.00022715)
$\text{GDP}_{pc} \times Weak_{noweight}$	-0.04302309*	-0.06543231**	-0.03504010*	-0.02895742	-0.04785540	-0.02475709
	(0.02202201)	(0.02927591)	(0.02076732)	(0.02225361)	(0.02949461)	(0.01982124)
$\text{GDP}^2_{pc} \times Weak_{noweight}$	0.00179473	0.00291153	0.00039388	0.00088749	0.00176080	-0.00013863
	(0.00198648)	(0.00264590)	(0.00212243)	(0.00200373)	(0.00264127)	(0.00193118)
$Weak_{noweight}$	0.03898439	0.06160696	0.18610820***	-0.02800673	-0.02299735	0.13448125***
	(0.05363940)	(0.07185015)	(0.04394072)	(0.05617752)	(0.07511852)	(0.04560524)
Constant	0.61441537***	0.50286605***	0.12846839***	0.64493595***	0.54146071***	0.15655408***
	(0.00803239)	(0.01072036)	(0.00537319)	(0.01312548)	(0.01792032)	(0.00868449)
Country FE	✓	✓	✓	✓	✓	✓
Year FE				✓	✓	✓
Observations	1736	1736	1736	1736	1736	1736
R^2	0.738	0.744	0.791	0.748	0.753	0.802

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A Appendix

Tab. A.1: Sample coverage

Albania (3), Algeria (28), Argentina (14), Bahamas (4), Bangladesh (28), Belgium (17), Belize (2), Benin (7), Bhutan (1), Bolivia (31), Botswana (12), Brazil (5), Bulgaria (11), Burkina Faso (10), Burundi (17), Cameroon (25), Central African Republic (16), Chile (38), China (22), Colombia (38), Congo, Rep. (14), Costa Rica (21), Cte d'Ivoire (21), Croatia (3), Czech Republic (2), Dominican Republic (23), Ecuador (37), Egypt, Arab Rep. (35), El Salvador (29), Eritrea (10), Estonia (2), Ethiopia (21), Fiji (25), Gabon (8), Gambia (8), Ghana (28), Guatemala (19), Guyana (1), Honduras (26), Hong Kong (13), Hungary (8), India (39), Indonesia (32), Iran, Islamic Rep. (36), Jamaica (28), Jordan (27), Kenya (39), Latvia (9), Lesotho (4), Luxembourg (1), Macao, China (19), Macedonia, FYR (7), Madagascar (22), Malawi (32), Malaysia (33), Malta (8), Mauritius (19), Mexico (31), Mongolia (6), Morocco (24), Namibia (1), Nepal (9), Nicaragua (21), Nigeria (28), Oman (9), Pakistan (30), Panama (37), Papua New Guinea (27), Paraguay (6), Peru (14), Philippines (34), Poland (11), Puerto Rico (14), Romania (12), Russian Federation (9), Saudi Arabia (1), Senegal (24), Sierra Leone (1), Slovak Republic (5), Slovenia (12), South Africa (30), Sri Lanka (28), Sudan (1), Swaziland (16), Syrian Arab Republic (35), Tanzania (9), Thailand (18), Togo (10), Tonga (1), Tunisia (28), Turkey (33), Uganda (1), Uruguay (31), Venezuela, RB (34), Vietnam (1), Yemen, Rep. (4), Zambia (18), Zimbabwe (34).

In parenthesis, the number of observations per country.

References

[1] Acemoglu, Daron, and Fabrizio Zilibotti, 1997. Was Prometheus Unbound by Chance ? Risk, Diversification, and Growth. Journal of Political Economy 105(4), pp. 709-51.

[2] Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar and Alireza Tahbaz-Salehi, 2012. The Network Origins of Aggregate Fluctuations. Econometrica 80(5), pp. 1977-2016, 09

[3] Cadot, Olivier, Celine Carrere and Vanessa Strauss-Khan, 2011. Export Diversification: what's behind the hump ? The Review of Economics and Statistics 93(2), pp. 590-605.

[4] Corden, Max, and Peter Neary, 1982. Booming Sector and De-industrialisation in a Small Open Economy. The Economic Journal 92 (December), pp. 825-48.

[5] Faini, Ricardo, 2004. Trade Liberalization in a Globalizing World. CEPR Discussion Paper #4665.

[6] Fuller, Wayne A. 1986. Measurement Error Models. John Wiley, New York.

[7] Gawande, Kishore, Pravin Krishna, and Marcelo Olarreaga, 2012. Lobbying Competition Over Trade Policy. International Economic Review, 53(1), pp. 115-132.

[8] Hirschman, Albert, 1958. The strategy of economic development. Yale University Press, New Haven.

[9] Imbs, Jean, and Romain Wacziarg, 2003. Stages of diversification. American Economic Review 93(1), pp. 63-86.

[10] Jones, Charles I., 2011. Intermediate Goods and Weak Links in the Theory of Economic Development. American Economic Journal: Macroeconomics, 3(2), pp. 1-28.

[11] Koren, Miklos, and Silvana Tenreyro, 2007. Volatility and development. Quarterly Journal of Economics 122(1), pp. 243-87.

[12] Kremer, Michael, 1993. The o-ring theory of economic development. Quarterly Journal of Economics 108(3), pp. 551-75.

[13] Lind, Jo Thori, and Halvor Mehlum, 2010. With or Without U ? The Appropriate Test for a U-Shaped Relationship. Oxford Bulletin of Economics and Statistics 72(1), pp. 109-18.

[14] Olarreaga, Marcelo, and Cristian Ugarte, 2012. Stages of development in Middle East and North Africa, Chp. 4 in Diop, Ndiame, Daniela Marotta, and Jaime de Melo eds., Natural Resource Abundance, Growth and Diversification in MENA, World Bank.

[15] Saint-Paul, Gilles, 1992. Technological Choice, Financial Markets and Economic Development. European Economic Review 36(4), pp. 763-81.

[16] Ugarte, Cristian, 2013. Inclusive Growth and the Impact of Intermediate Goods' Productivity on Economic Development, PhD thesis, University of Geneva, Switzerland.